中西方文化差异对虚拟人道德责任判断的影响*

闫 霄　莫田甜　周欣悦
(浙江大学管理学院，杭州 310058)

摘要：近年来，社交网络上涌现出大量具有真人特征的虚拟人物，他们以第一人称视角叙事以获得影响力，然而其背后的道德伦理问题也逐渐浮现。当社交媒体虚拟人犯错时，不同文化背景的人们对其道德判断会有何不同？本文基于心智感知理论，通过五个主要实验探索中西方文化差异对虚拟人道德责任判断的影响和机制。研究表明，在被告知虚拟人的不道德行为后，比起西方文化，中国文化影响下的人们认为虚拟人需要承担更大的道德责任；但这种文化差异没有出现在对真人的道德责任判断上(实验 1a~1c)。并且不论虚拟人是由真实人类还是人工智能驱动，这种文化差异都会存在(实验 1c)。这种文化差异通过感知心智能力来中介(实验 2)。具体而言，比起西方文化，中国文化影响下的人们认为虚拟人的心智能力(特别是感知力)更高，因此需要承担更大的道德责任。此外，对虚拟人更大的道德责任判断会导致中国文化(相比西方文化)影响下的人们更倾向于对虚拟人施加道德惩罚(实验 3)。本文通过实证研究将道德责任判断和心智感知的对象拓展到虚拟人上，并揭示了中西方文化差异及其后续影响。

关键词：虚拟人，道德责任，文化差异，心智能力

1 问题提出

虚拟人是指通过计算机技术生成的具有真人外观、特征和行为的虚拟形象(Sung et al., 2022)。随着元宇宙概念的兴起，越来越多的虚拟人以各种各样的形式出现在全球各地，主要包括活跃在社交媒体平台上的虚拟博主和虚拟偶像，例如我国的洛天依、美国的 Lil Miquela、日本的 Imma 等。它们利用计算机图像软件制作的数字形象，以第一人称视角看待世界，并在社交媒体平台上占据一席之地(Arsenyan & Mirowska, 2021)。从 2019 年科技部

* 本文系国家杰出青年科学基金(71925005)，浙江省哲学社会科学领军人才培育专项课题(21YJRC01ZD)的研究成果之一。

等部门发布的《关于促进文化和科技深度融合的指导意见》到 2022 年国务院发布的《“十四五”数字经济发展规划》，国家明确鼓励虚拟数字技术的研发与实践。在政策的推动下，2021 年虚拟人相关企业融资达 2843 起，融资金额高达 2540 亿元[1]。然而，在社会关注和资本加持的驱动下，相关的法律和道德伦理问题也不断产生。

其中一个重要问题就是，人们会对虚拟人犯错做出怎样的道德责任判断？现实案例表明，不同文化背景下，人们的态度也有所不同。在西方国家，虚拟人 Bermuda 非法盗用了一位巴西裔美国网红 Miquela 的 Instagram 账号并且擅自删除账号内容，但其做法非但没有受到责备，还获得了外界的认可(Feller, 2018)。美国著名杂志《连线》的一名资深编辑曾就《第二人生》游戏中发生的虚拟人强奸事件做出评价，反对了虚拟犯罪的观点(Lynn, 2007)。研究还发现，在欧美等国外的政治大选中存在社交虚拟人在线诋毁候选人的情况(Kollanyi et al., 2016)，他们会传播未经证实的健康声明，甚至对公众健康造成危害(Allem et al., 2020)，但并未得到有效治理和管控。然而在我国，人们对虚拟人的不道德行为并不宽容。例如，哔哩哔哩平台打造的虚拟女团“四禧丸子”刚一出道就被指认作品抄袭，引起了网民的高度不满，最终导致官方发文道歉并下架作品。可见，虚拟人会在网络上进行侵权、抄袭、盗用、诋毁等不道德行为，并且不同文化背景下的人们可能态度也不一致。那么，相比于真实人类在社交网络上进行同样的不道德行为，人们会如何看待虚拟人的不道德行为？例如，人们在多大程度上会认为虚拟人应该为之负责？这种判断在不同文化下存在什么差异？现有研究尚未探索这些研究问题。本文将通过五个主要实验考察中西方文化差异对虚拟人道德责任判断的影响和机制，对虚拟实体道德判断的相关研究进行延伸，丰富文化差异、心智感知等相关文献，并为虚拟人设计、运营和在道德伦理上的治理问题提供一定的实践启示。

1.1 中西方文化差异对道德判断的影响

文化会对人们的心理与行为产生重要影响(侯玉波, 朱滢, 2002)。文化泛指社会成员之间所共有的价值观、规范、思维方式和行为方式等(Morling, 2016; Na et al., 2010; 黄梓航 等, 2018)。先前的研究发现，东西方文化差异会影响人们的道德判断。西方文化强调以个人自由、权利、公正、关爱和宽恕等为导向的道德观，而东方文化强调以集体和个人责任为导向的道德观(彭凯平 等, 2011; 王恩界, 乐国安, 2006)。Miller 和 Bersoff (1992)的研究指出，西方人更强调公正伦理，东方人更强调责任伦理。个人主义/集体主义文化也会影响人们的道德决策，因为它们涉及到个人利益还是集体利益优先的信念(Oyserman et al., 2002)。基于个人主义/集体主义的文化心理学理论，本文关注到了中西方文化差异对虚拟人道德责任判断

[1] 数据源自速途元宇宙研究院联合北京城市副中心应用场景产业联盟发布《2022 虚拟人产业研究报告》

的影响。与个人主义文化(如美国)相比，集体主义文化(如中国)下的人们会有更高的换位思考能力(Wu & Keysar, 2007)和拟人化倾向(Letheren et al., 2016)。相比于西方文化，中国文化影响下的人们是否会对虚拟人的心智能力评价更高，进而认为虚拟人应该为不道德行为承担更大的道德责任？这是本文想要回答的问题。

1.2 对虚拟人的道德责任判断

在以往文献中，大多数学者认为虚拟人并非真实存在的机器人或是单纯的电脑程序，他们将虚拟人定义为计算机生成的拟人化图像(Arsenyan & Mirowska, 2021; Kim & Jo, 2022; Sung et al., 2022)，并在人工智能技术的加持下能够用语言、姿势，甚至表情来和人们进行面对面的交流互动(Kim & Jo, 2022; Volante et al., 2016)。目前有关虚拟人的研究集中在探讨虚拟人与真人在社交媒体上的互动(Arsenyan & Mirowska, 2021)、虚拟偶像特征及代言效果(Sands et al., 2022)。本文关注于社交媒体上的虚拟人，他们是利用计算机图像软件制作的数字形象，以第一人称视角看待世界，并在媒体平台上占据一席之地。相比于任务导向的人工智能机器人和算法软件，社交媒体平台上的虚拟人以获取影响力为目的进行分享与互动，与真实人类具有更相似的行为，更容易出现不道德行为。尽管虚拟人的不道德行为被频繁曝光，但尚未有研究直接探讨人们对虚拟人的道德责任判断，大多数研究聚焦人工智能、算法等虚拟实体。由于虚拟人的产生同样依托人工智能技术，我们可以借鉴相关虚拟实体的研究。

道德责任的归属问题是道德哲学的基本问题，可以追溯到亚里士多德提出的理论，他认为一个实体应该同时满足自由意志和认知条件，才能作为道德主体承担道德责任(Aristotle, 1999)。自由意志条件讨论了代理人是否有可能采取不道德行动，认知条件考察了代理人是否能够为行为后果承担道德责任(Clarke, 1992; Constantinescu et al., 2022; Zimmerman, 1997)。在讨论人工智能等虚拟实体是否能够承担道德责任时，许多学者也提到了这两个基本条件。相关文献指出，虚拟实体虽然具有一定程度的自主能力，但缺乏认知能力，仍不能作为责任主体承担责任(Hakli & Mäkelä, 2019; Parthemore & Whitby, 2014)。如果虚拟实体进行了不道德行为，其背后真实存在的人(如设计师或制造商等)应该对其造成的任何后果承担责任(Bryson et al., 2017; Champagne & Tonkens, 2015; Constantinescu et al., 2022)。

然而，单独从两个责任主体条件的角度去分析道德责任的归属是存在缺陷的(Shoemaker, 2011)。Coeckelbergh(2020)强调了责任客体的重要性，道德责任问题不仅仅涵盖做什么和知道自己在做什么，还应包括谁对某人或某事负责的问题，即可回答性，这种行为双方的对应关系能为道德责任的归属问题提供更多依据。然而对于虚拟实体来说，很多行为后果是难以解释和回答的，因此客观上难以理清道德责任的归属，但主观上人们依然会对这些虚拟实体

产生道德判断(Young & Monroe, 2019; Malle et al., 2015; 褚华东 等, 2019)。基于上述分析，本文预测，人们在对虚拟人进行道德判断时也会从责任主体和责任客体两方面因素去归因。

1.3 心智感知理论

心智是指思考、感受和有意识的行为能力(Tharp et al., 2017)。心智感知，即在其他实体(如技术设备、精神代理和非人类动物)中感知心智的存在，有助于理解和预测他人的行为和意图(Johnson & Barrett, 2003; Waytz et al., 2010)。根据心智感知理论(Mind Perception Theory)，心智能力(mental capacity)涉及认知的自主力(agency)和情绪的感知力(experience)，包括计划、自我控制和情绪感受等(Gray et al., 2007; Gray et al., 2011; Gray & Wegner, 2009)。将认知自主力归因于一个实体意味着观察者认为该实体能够作为道德主体，像成年人一样行动、计划、实施自我控制等；将情绪感知力归因于一个实体表明观察者相信该实体能够作为道德客体，存在情绪感知能力(Gray et al., 2007)。两者可以被单独感知。例如，婴儿被认为具有高情绪感知力和低认知自主力，上帝被认为具有高认知自主力和低情绪感知力(Gray et al., 2007)。

心智能力会影响对实体的道德判断，在二元道德概念中发挥作用(Gray et al., 2012; Ward et al., 2013)。具有认知自主力的实体能自己做出决定并有意识地采取行动，当行为出现错误时，他们被认为应该为此负责(Gray et al., 2012; Sullivan & Wamba, 2022)。正常的成年人具备认知自主力，可以作为道德主体承担道德责任。未成年人不具备认知自主力，在做出错误道德行为时，会受到更轻微的惩罚(Gray & Wegner, 2009)。虽然认知自主力对道德责任的判断至关重要，但情绪感知力是区分人类和其他实体的本质特征，通常被认为具有更强大的解释力(Gray & Wegner, 2012)。心理学研究表明，情绪感知力在道德决策中发挥重要作用(Greene et al., 2001; Haidt et al., 1993)。感受他人的痛苦并表现出同理心是道德判断的核心要素之一(Bigman & Gray, 2018)。情绪感知力也对道德主体的判断起到重要作用(Bigman & Gray, 2018)。例如，Sulliva 和 Wamba (2022)发现，当人们认为人工智能故意伤害人类(相比于非人类)时，会将责任归咎于人工智能，这是由心智能力中的情绪感知力来中介的。基于心智感知理论，本文认为，当社交媒体上的虚拟人犯错时，如果赋予其类似正常成年人的高心智能力，人们可能会认为虚拟人需要承担和真人相似的道德责任。

1.4 心智感知及道德责任判断的文化差异

已有文献认为，文化差异是影响心智感知的重要因素(Dietze & Knowles, 2021)。一方面，文化差异会影响人们的换位思考能力。例如，比起个人主义文化(如美国)，集体主义文化(如中国)下的人们会表现出更高的换位思考能力(Wu & Keysar, 2007)。比起英国人，中国人对人

工智能机器人社会合作属性的感知更高(Dang & Liu, 2022)。因此，比起西方人，中国人更可能将自己的情感和能力投射到别人身上，可能会对虚拟人的心智能力评价更高。另一方面，文化差异会影响人们的拟人化倾向。例如，Letheren 等(2016)发现，东亚人比白种人具有更高的拟人化倾向。西方文化的基督教和伊斯兰教认为神更不具有固定形状的身体，而东方文化的佛教和印度教认为神具有有形的身体(Fuller, 2004; Ohnuma, 2007; Samuel, 1989; Willard & McNamara, 2019)。东方人还会给神赋予人类特有的能力(McGuire, 2018)。而拟人化倾向会影响心智能力的感知，拟人化倾向越高的被试对机器模型 Pleo 心智能力的感知越高(Eyssel & Pfundmair, 2015)。因此，东方人比西方人更可能对非人类实体进行拟人化，可能会更加认为虚拟人具备心智能力。综上，本文认为，比起西方文化，中国文化影响下的人们可能会对虚拟人的心智能力评价更高。

心智感知理论将心智能力划分为认知自主力和情绪感知力两个维度(Gray et al., 2007)。对一个实体赋予更高的认知自主力意味着该实体能够作为道德主体做出道德决策并为行为负责(Gray et al., 2012; Gray & Wegner, 2009; Himma, 2009)。研究发现，情绪感知力对道德主体的判断同样至关重要(Greene et al., 2001; Sullivan & Wamba, 2022)。因此，对一个实体心智能力的评价可能会影响人们对其道德责任的判断。然而，目前对虚拟实体道德责任判断的文献存在分歧。部分学者认为虚拟实体不能承担道德责任(Hakli & Mäkelä, 2019; Parthemore & Whitby, 2014)。也有研究表明，人们会将道德责任归因于虚拟实体(Awad et al., 2020; Malle et al., 2015; Young & Monroe, 2019; 褚华东 等, 2019)。为了得出一致结论，需要深入探索人们对虚拟实体做出道德判断背后的原因。结合心智感知在道德责任判断中的重要作用，本文推断，中西方文化差异能通过影响心智能力的评价，来影响人们对虚拟实体的道德责任判断。

目前针对虚拟实体道德责任的研究大多针对人工智能，并采用道德两难问题的研究范式(Awad et al., 2020; Malle et al., 2015; Young & Monroe, 2019; 褚华东 等, 2019)。而传统道德两难问题(如电车实验)的道德情景将理性与直觉剥离开，被认为是非典型的(Schein & Gray, 2018)。随着越来越多的虚拟人出现在人们的日常生活中，以虚拟实体为主体的道德问题不再局限于道德困境，取而代之的是更多类人的不道德行为，如侵权、抄袭等。因此，本文聚焦社交媒体上的虚拟人贴近现实的道德情景，进行了实证研究，并提出以下研究假设：

假设 1：当看到虚拟人进行不道德行为后，中国文化(比西方文化)影响下的人们认为虚拟人需要承担更大的道德责任。

假设 2：中国文化(比西方文化)影响下的人们认为虚拟人具备更高的心智能力，这种更高的心智能力归因导致他们更加认为虚拟人应该为不道德行为承担道德责任。

1.5 研究概览

综上所述，本研究基于心智感知理论，通过五个实验考察了中西方文化差异对虚拟人道德责任判断的影响。实验 1a 发现，中国文化(比西方文化)影响下的人们认为虚拟人犯错需要承担更大的道德责任。实验 1b 换用不同的不道德情景和道德责任的测量复制了实验 1a 的结果。实验 1c 操纵了虚拟人背后主体的类型，无论虚拟人由真实人类还是人工智能驱动，这种中西方文化差异的效应依然存在。实验 2 考察了感知心智能力的中介机制，中国文化影响下的人们对虚拟人更高的心智能力感知导致他们认为虚拟人犯错应该承担更大的道德责任。实验 3 进一步探讨了这种文化差异对人们后续行为(道德惩罚)的影响。

2 实验 1a：网络暴力情境

实验 1a 的目的是考察文化背景对人们道德责任判断的影响。本实验采用 2(真人 vs.虚拟人) × 2 (中国文化 vs.西方文化)被试间的实验设计，考察相比于西方文化，中国文化影响下的人们是否会认为虚拟人在犯错之后应该承担更大的道德责任。

2.1 实验设计与样本

实验采用 G*Power 3.1 软件(Faul et al., 2007)计算实验所需样本量。对于本实验适用的双因素方差分析，当显著性水平α为 0.05 且效应量为中等效应时(f = 0.2 到 0.25 之间)，要达到 95%统计检验力所需要的样本量在 279 到 434 之间。为保证实验具有足够的样本量且为统一标准，本研究的所有实验按照每个条件下 100 名被试的标准招募，且采用原始数据进行分析。但由于线上平台存在同时作答以及未完成作答等情况，会出现多于或少于既定被试数量 1~2 个的情况。

实验 1a 通过国外问卷平台 Prolific 和国内问卷平台 Credamo 分别邀请了 200 名美国白人被试(57.50%女性; $M_{年龄}$ = 29.49 岁, $SD_{年龄}$ = 6.26 岁)和 200 名中国被试(75.00%女性; $M_{年龄}$ = 25.16 岁, $SD_{年龄}$ = 5.21 岁)参与研究，并以小额现金作为实验报酬。两种文化背景的被试被随机分配到真人组和虚拟人组，每组各 100 人。

2.2 实验流程

在真人组中，美国(中国)的被试会看到一位 Twitter(微博)博主 Rico 的主页(如图 1 所示)和简短的文字介绍；在虚拟人组中，被试看到的则是虚拟博主 Rico 的主页(如图 2 所示)和相同的文字介绍以及对虚拟博主的定义(虚拟博主是利用计算机图像软件制作的数字形象，他们以第一人称视角看待世界，并在媒体平台上占据一席之地)。事实上，Rico 是虚构的博主，

其主页头像采用免费网站(https://generated.photos)人工智能自动生成的真实人形象，对应的虚拟人头像则是根据真人形象绘制。其中，文字介绍内容为:

“Rico 是微博/Twitter 上的一个(虚拟)博主。她经常活跃在社交媒体上分享她的穿搭和生活。”

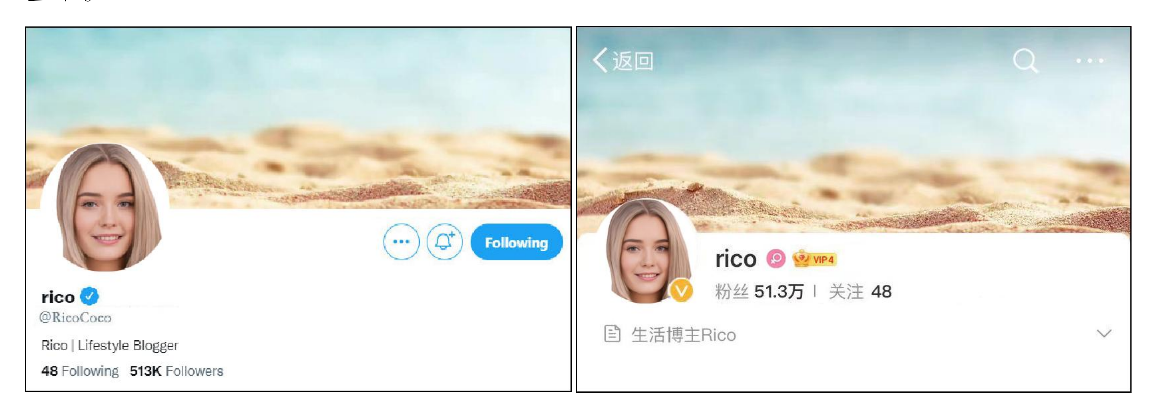


图 1 实验 1a 中真人组 Twitter 主页(左)和微博主页(右)

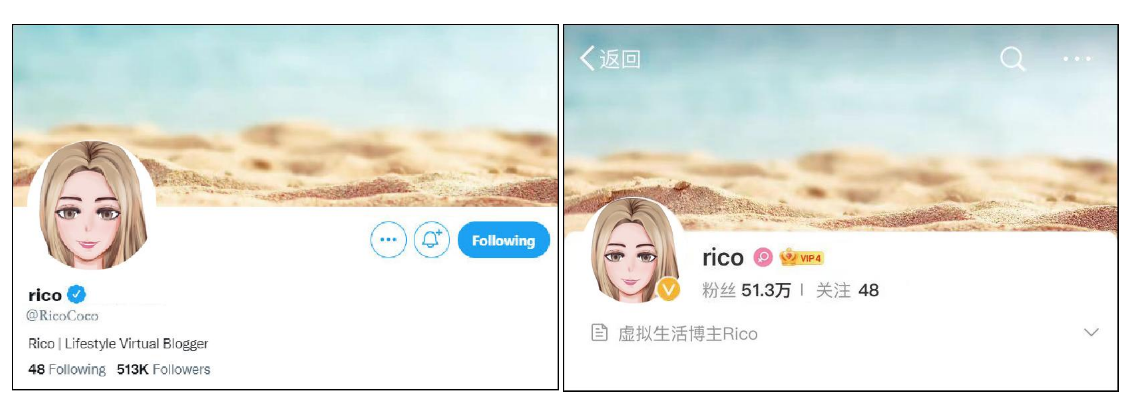


图 2 实验 1a 中虚拟人组 Twitter 主页(左)和微博主页(右)

接下来，被试被告知，“最近，Rico 在微博上发帖曝光某一位网民的隐私行为，致使该网民遭到网络暴力。”为了测量道德责任判断，被试回答了“你认为对于这起网络暴力事件，Rico 应该承担多少责任(7 分量表, 1 = 没有责任, 7 = 很大责任)”以及“你认为对于这起网络暴力事件，Rico 应该承担多少指责(7 分量表, 1 = 没有指责, 7 = 很多指责)”，测项语句的构成和选取在借鉴前人研究的基础上有所调整(Gray & Wegner, 2009)。由于道德责任判断的两个题目高度相关($r = 0.76$, $p < 0.001$)，因此取二者的均值作为道德责任判断。最后，所有被试报告了性别和年龄。本研究中所有实验的完整操纵和测量材料见附录一。

2.3 数据分析与结果

以文化类型(美国编码为 0，中国编码为 1)和博主类型(真人编码为 0，虚拟人编码为 1)分别作为自变量，道德责任判断作为因变量进行方差分析。结果显示，文化类型的主效应显著($F(1, 396) = 4.41$, $p = 0.036$, $\eta_p^2 = 0.011$)，博主类型的主效应也显著($F(1, 396) = 7.23$, $p = 0.007$, $\eta_p^2 = 0.018$)，文化类型和博主类型对道德责任判断的交互效应显著($F(1, 396) = 5.83$, p

= 0.016, η_p^2 = 0.015)，如图 3 所示。具体而言，比起美国人，中国人认为虚拟人应该承担更大的道德责任($M_{中国}$ = 5.48, SD = 1.07, 95%CI = [5.23, 5.73]; $M_{美国}$ = 4.91, SD = 1.68, 95%CI = [4.66, 5.16]; F(1, 396) = 10.19, p = 0.002, η_p^2 = 0.025)；而对真人道德责任的判断没有显著的差异(F(1, 396) = 0.05, p = 0.824)。此外，中国人对真人和虚拟人的道德责任判断不存在显著差异(F(1, 396) = 0.04, p = 0.846)。以上数据结果表明，当看到虚拟人进行网络暴力的不道德行为后，中国文化(vs.西方文化)影响下的人们认为虚拟人需要承担更大的道德责任，从而验证了假设 1。本实验和后续实验在控制性别和年龄等变量后均不改变结果的显著性和方向，详细数据结果见附录四。

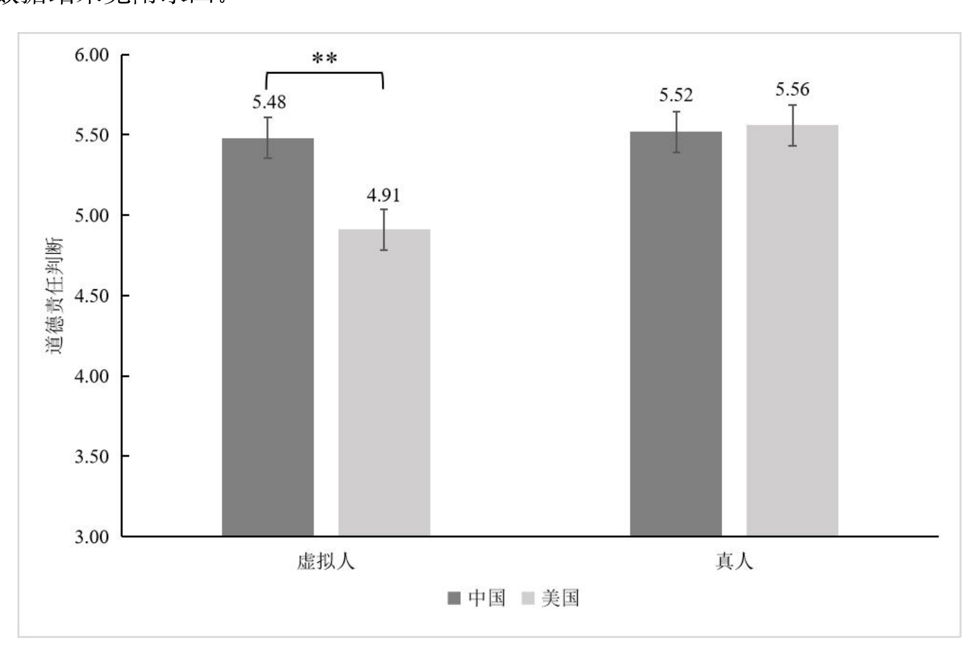


图 3 实验 1a 中不同文化类型下对虚拟人和真人的道德责任判断评分(Error bars: 95%CI; ** p < 0.01)

2.4 讨论

实验 1a 采用曝光隐私引发网络暴力的不道德行为情景来检验中西方文化对虚拟人作为道德主体在道德责任判断上的差异。结果显示，当人们看到虚拟人的不道德行为后，相比于西方文化，中国文化影响下的人们认为其需要承担更大的道德责任。这一结果从文化差异的角度拓展了人们对虚拟实体道德责任归因的文献(Awad et al., 2020; Malle et al., 2015; Young & Monroe, 2019; 褚华东 等, 2019)，表明不同文化背景下的人们对虚拟人道德责任的判断存在差异。

3 实验 1b：偷税漏税情境

实验 1b 的目的是采用另一种不道德情景和另一种道德责任的测量方法，并招募另一个西方国家的被试来复制实验 1a 的结果。我们预计，相比于西方文化，中国文化影响下的人

们认为虚拟人偷税漏税需要承担更大的道德责任。

3.1 实验设计与样本

实验 1b 也采用 2(真人 vs.虚拟人) × 2 (中国文化 vs.西方文化)被试间的实验设计。我们通过国外问卷平台 Prolific 和国内问卷平台 Credamo 分别邀请了 199 名英国白人被试(68.84%女性; $M_{年龄}$ = 29.87 岁, $SD_{年龄}$ = 6.06 岁)和 200 名中国被试(64.50%女性; $M_{年龄}$ = 26.64 岁, $SD_{年龄}$ = 5.52 岁)参与研究，并以小额现金作为实验报酬。两种文化背景的被试被随机分配到真人组和虚拟人组。在英国被试中，真人组有 99 人，虚拟人组有 100 人；在中国被试中，真人组和虚拟人组各 100 人。

3.2 实验流程

与实验 1a 的操纵方式和内容类似，在真人组中，中国(英国)的被试会看到博主 Rico 的微博(Twitter)主页及描述；在虚拟人组中，中国(英国)的被试会看到虚拟博主 Rico 的微博(Twitter)主页及描述。接下来，被试被告知，“最近，Rico 被曝出在直播带货的收入上存在偷税漏税的行为。现在，她正在被调查中。”接着，本实验用四个条目的道德责任量表来测量被试对 Rico 道德责任的判断(7 分量表, 1 = 完全不同意, 7 = 完全同意; Cameron et al., 2010; Cronbach’s α = 0.81)。其中两个条目为“Rico 在道德上应该为她偷税漏税的行为负责”和“Rico 应该为她偷税漏税的行为受到处罚”，得分越高表明被试对 Rico 道德责任的判断越高；另外两个条目为“偷税漏税的行为不应该归咎于 Rico”和“偷税漏税的责任不应该由 Rico 承担”，得分越高表明被试对 Rico 道德责任的判断越低。我们将前两个条目的得分和后两个条目的反向得分相加平均后，得到了道德责任判断的测量得分。随后，被试回答了对博主类型操纵检验的问题：“就你而言，Rico 是？(0 = 真人博主, 1 = 虚拟博主)”最后，被试报告了性别和年龄。

3.3 数据分析与结果

操纵检验 以博主类型(真人编码为 0，虚拟人编码为 1)为自变量，文化类型(英国编码为 0，中国编码为 1)为调节变量，博主类型的操纵检验(0 = 真人博主, 1 = 虚拟博主)为因变量，在样本选择为 5000 次、95%的置信区间下采用 Bootstrapping (PROCESS Model 1; Hayes, 2015)进行回归分析。结果显示，文化类型的主效应不显著($\beta = -0.32$, $SE = 0.56$, $p = 0.567$, 95%CI = [−1.42, 0.78])，博主类型和文化类型的交互效应也不显著($\beta = 1.02$, $SE = 1.35$, $p = 0.450$, 95%CI = [−1.63, 3.58])，仅在博主类型的主效应上存在显著差异($\beta = 6.32$, $SE = 0.80$, $p < 0.001$, 95%CI = [4.75, 7.90])。具体而言，在虚拟人组中，分别有 98.00%的英国被试和 99.00%的中国被试认为 Rico 是虚拟博主；在真人组中，分别有 91.92%的英国被试和 94%的中国被试认

为 Rico 是真人博主。这说明关于博主类型的操纵是有效的，且不存在文化上的差异。此外，排除不符合操纵检验要求的被试并不会影响现有结果的方向和显著性。

道德责任判断 以文化类型(英国编码为 0，中国编码为 1)和博主类型(真人编码为 0，虚拟人编码为 1)为自变量，道德责任判断为因变量进行方差分析。结果显示，文化类型的主效应显著($F(1, 395) = 8.50$, $p = 0.004$, $\eta_p^2 = 0.021$)，博主类型的主效应显著($F(1, 395) = 28.56$, $p < 0.001$, $\eta_p^2 = 0.067$)，文化类型和博主类型对道德责任判断的交互效应显著($F(1, 395) = 9.40$, $p = 0.002$, $\eta_p^2 = 0.023$)，如图 4 所示。具体而言，比起英国人，中国人认为虚拟人应该承担更大的道德责任($M_{中国} = 5.36$, $SD = 1.19$, 95%CI = [5.12, 5.61]; $M_{英国} = 4.61$, $SD = 1.66$, 95%CI = [4.36, 4.86]; $F(1, 395) = 17.93$, $p < 0.001$, $\eta_p^2 = 0.043$)；但对真人的道德责任判断没有显著差异($F(1, 395) = 0.01$, $p = 0.915$)。此外，中国人对真人和虚拟人的道德责任判断不存在显著差异($F(1, 395) = 2.60$, $p = 0.108$)。结果复制了实验 1a 的发现，再一次验证了假设 1。

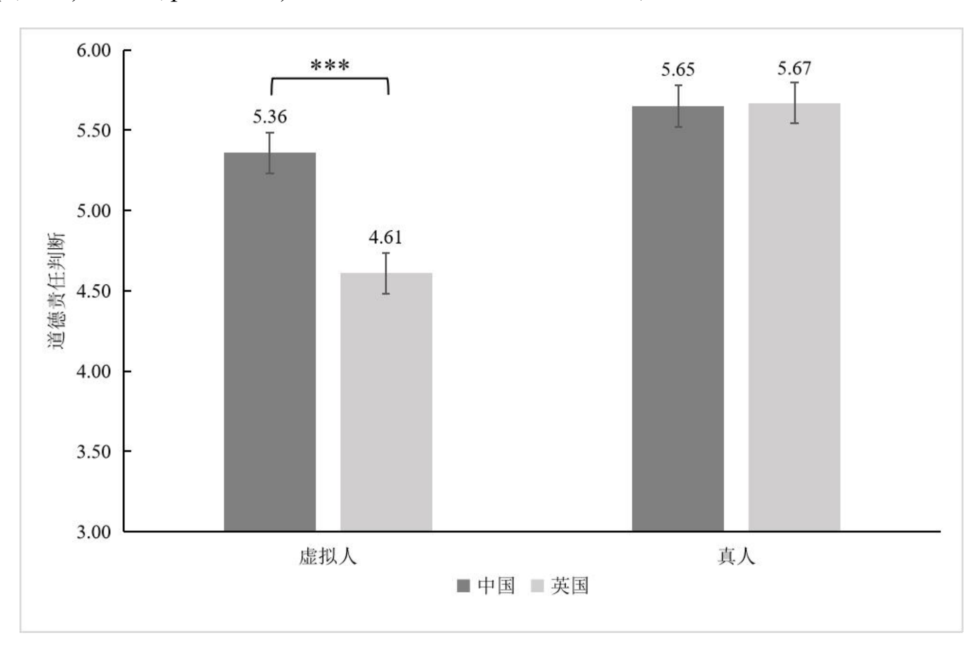


图 4 实验 1b 中不同文化类型下对虚拟人和真人的道德责任判断评分(Error bars: 95%CI; *** $p < 0.001$)

3.4 讨论

实验 1b 通过新的不道德行为情景(偷税漏税)和道德责任测量条目(Cameron et al., 2010)表明，比起西方文化，中国文化影响下的人们认为虚拟人在出现偷税漏税的行为后需要承担更大的道德责任。实验 1b 复制了实验 1a 的结果，扩展了与虚拟实体道德责任相关的研究(Awad et al., 2020; Malle et al., 2015; Young & Monroe, 2019; 褚华东 等, 2019)。在实验 1c 中，我们将考察是否不同类型的虚拟人会影响人们对他们的道德责任判断。

4 实验 1c：虚拟人主体类型的影响

当面对社交媒体虚拟人时, 人们可能会联想到这些虚拟人背后的主体究竟是真实人类还是人工智能技术，从而存在不同的判断。因此，实验 1c 考察是否不同的虚拟人主体类型会影响人们对虚拟人的道德责任判断。

4.1 实验设计与样本

实验 1c 采用 3(真人 vs.虚拟人工智能 vs.虚拟真人) × 2 (中国文化 vs.西方文化)被试间的实验设计，通过国外问卷平台 Prolific 和国内问卷平台 Credamo 分别招募了 300 名英国白人被试(65.00%女性; $M_{年龄}$ = 30.82 岁, $SD_{年龄}$ = 5.37 岁)和 300 名中国被试(68.33%女性; $M_{年龄}$ = 26.78 岁, $SD_{年龄}$ = 5.84 岁)参与研究，并以小额现金作为实验报酬。两种文化背景的被试被随机分配到真人组、虚拟人工智能组或虚拟真人组。在英国被试中，真人组有 99 人，虚拟人工智能组有 100 人，虚拟真人组有 101 人；在中国被试中，每组各有 100 人。

4.2 实验流程

博主类型的操纵方法与前面实验的操纵方式类似。不同的是在真人组，被试看到的是有关真人博主的描述："Rico 是微博(Twitter)上的一个博主"；在虚拟人工智能组，被试看到的是有关虚拟博主的描述："Rico 是微博(Twitter)上的一个由人工智能和算法驱动的虚拟博主"；在虚拟真人组，人们看到的是有关虚拟博主的描述："Rico 是微博(Twitter)上的一个由真实人类驱动的虚拟博主"。接着，被试被告知，"最近，Rico 在微博上公开发表的一张图片被指控侵犯了某位摄影师的作品版权。事后，Rico 删除了这张图片并公开道歉。"随后, 被试报告了对 Rico 的道德责任判断, 测量条目与实验 1a 用到的测量相同(r = 0.71, p < 0.001)。另外，虚拟人组的被试还回答了操纵检验的问题，"就你而言，驱动 Rico 的背后主体是(0 = 真实人类，1 = 人工智能)"。考虑到社会经济地位可能会影响个体的道德判断(Sui & Mo, 2022)，所有被试在最后报告了主观社会经济地位(10 点量表; Adler et al., 2000)、受教育程度(5 点量表)、性别和年龄。

4.3 数据分析与结果

操纵检验 在虚拟组的被试中，以博主类型(虚拟真人编码为 0，虚拟人工智能编码为 1)为自变量，文化类型(英国编码为 0，中国编码为 1)为调节变量，博主类型的操纵检验(0 = 真实人类，1 = 人工智能)为因变量，在样本选择为 5000 次、95%的置信区间下采用 Bootstrapping (PROCESS Model 1; Hayes, 2015)进行回归分析。结果显示，文化类型的主效应不显著(β = 0.08, SE = 0.36, p = 0.829, 95%CI = [−0.64, 0.79])，博主类型和文化类型的交互效应也不显著

($\beta = -0.71$, $SE = 0.68$, $p = 0.298$, 95%CI = [−2.05, 0.63])，仅在博主类型的主效应上存在显著差异($\beta = 4.47$, $SE = .53$, $p < 0.001$, 95%CI = [3.44, 5.51])。具体而言，在虚拟人工智能组中，分别有 95.00%的英国被试和 91.00%的中国被试认为 Rico 由人工智能驱动；在虚拟真人组中，分别有 82.18%的英国被试和 81.00%的中国被试认为 Rico 由真人驱动。这说明关于虚拟人主体类型的操纵是有效的，且不存在文化上的差异。此外，排除不符合操纵检验要求的被试并不会影响现有结果的方向和显著性。

道德责任判断 以文化类型(美国编码为 0，中国编码为 1)和博主类型(真人编码为-1，虚拟真人编码为 0，虚拟人工智能编码为 1)分别作为自变量，道德责任判断作为因变量进行方差分析。结果显示，文化类型的主效应显著($F(1, 594) = 95.39$, $p < 0.001$, $\eta_p^2 = 0.138$)，博主类型的主效应也显著($F(2, 594) = 21.63$, $p < 0.001$, $\eta_p^2 = 0.068$)，文化类型和博主类型对道德责任判断的交互效应显著($F(2, 594) = 12.02$, $p < 0.001$, $\eta_p^2 = 0.039$)，如图 6 所示。具体而言，比起英国人，中国人认为无论虚拟人背后的主体是真实人类($M_{中国} = 5.41$, $SD = 1.26$, 95%CI = [5.14, 5.68]; $M_{英国} = 3.78$, $SD = 1.50$, 95%CI = [3.51, 4.05]; $F(1, 594) = 70.56$, $p < 0.001$, $\eta_p^2 = 0.106$)还是人工智能($M_{中国} = 5.13$, $SD = 1.27$, 95%CI = [4.85, 5.40]; $M_{英国} = 3.80$, $SD = 1.57$, 95%CI = [3.53, 4.07]; $F(1, 594) = 46.24$, $p < 0.001$, $\eta_p^2 = 0.072$)，都需要承担更大的道德责任；但对真人道德责任的判断没有显著的文化差异($F(1, 594) = 3.01$, $p = 0.084$)。此外，中国人对真人和两种类型的虚拟人的道德责任判断也没有显著差异($F(2, 594) = 1.83$, $p = 0.162$)，且两两比较的结果都不存在显著差异($ps > 0.203$)。以上结果说明，在得知虚拟人侵犯版权后，不论虚拟人背后的主体是真实人类还是人工智能，中国文化(相比于西方文化)影响下的人们都认为虚拟人需要承担更大的道德责任。结果进一步为假设 1 提供了支持。

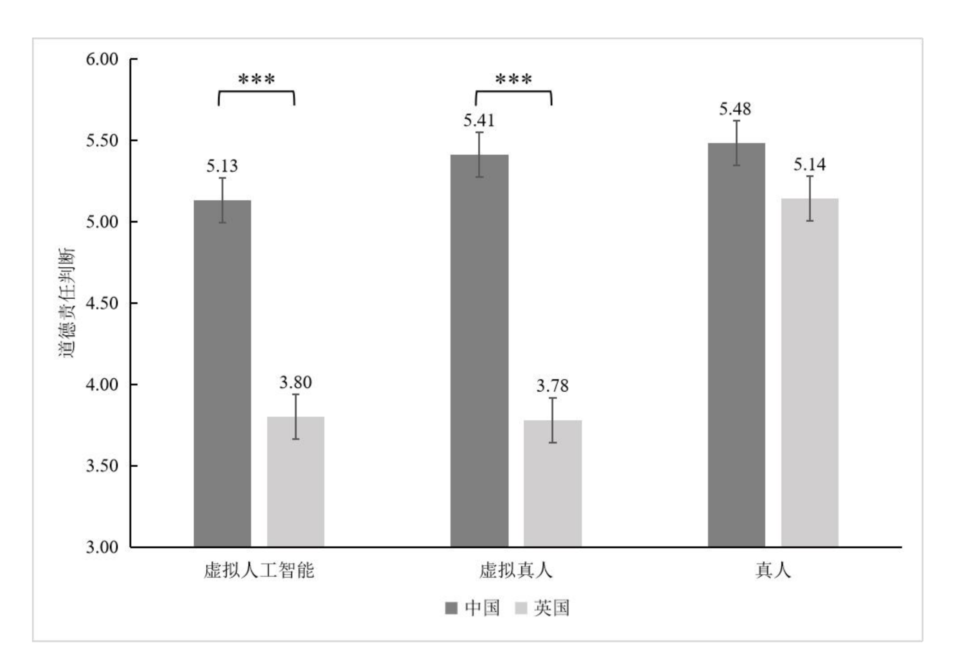


图 6 实验 1c 中不同文化类型下对两种虚拟人和真人的道德责任判断评分(Error bars: 95%CI; *** $p < 0.001$)

4.4 讨论

实验 1c 通过操纵虚拟人背后的主体类型，分别在真人组、虚拟真人组和虚拟人工智能之间比较中西方文化差异。结果发现，无论虚拟人背后的主体是真实人类还是人工智能，相比于西方文化，中国文化影响下的人们都会赋予其更大的道德责任，而在真人组不存在这种文化差异。这说明，道德责任判断的中西方文化差异同样存在于由真实人类和人工智能驱动的虚拟人当中。以往有研究认为，在对虚拟实体做出责任判断时会考虑其背后真实主体(Bryson et al., 2017; Champagne & Tonkens, 2015; Constantinescu et al., 2022)，但本实验结果发现虚拟人背后的主体类型并不会影响对其道德责任判断的文化差异。在实验 2 中，我们将考察对虚拟人感知心智能力的中介机制。

5 实验 2：感知心智能力的中介效应

实验 2 有两个目的。首先，实验 2 考察了中西方文化对虚拟人道德责任判断的差异是否源自于感知心智能力的差异。我们认为，相比于西方文化，中国文化影响下的人们认为虚拟人的心智能力更高，因此虚拟人在犯错后需要承担更大的道德责任。其次，由于该效应的产生可能是因为中国人比西方人更熟悉虚拟人，或者中国人比西方人认为不道德情景中的事件(如偷税漏税)更严重而导致的文化差异，与我们假设的机制无关。因此，实验 2 通过测量对虚拟人熟悉度和事件严重程度来排除这两种替代性解释。

在正式实验前，我们进行了一项预实验，通过国外问卷平台 Prolific 和国内问卷平台 Credamo 各招募了 199 名英国白人被试(80.40%女性; $M_{年龄}$ = 30.62 岁, $SD_{年龄}$ = 5.80 岁)和 200

名中国被试(77.50%女性; $M_{年龄}$ = 30.89 岁, $SD_{年龄}$ = 4.95 岁)参与研究。被试被随机分配到真人组或虚拟人组，让他们评价 Rico 的心智能力(7 分量表, 1 = 没有能力, 7 = 很有能力; Gray et al., 2011; Cronbach's α = 0.93)。结果发现，文化类型和博主类型对心智能力评价的交互效应显著($F(1, 395) = 38.00$, $p < 0.001$, $\eta_p^2 = 0.088$)。比起英国人，中国人认为虚拟人的心智能力更高($M_{中国}$ = 4.57, SD = 1.08; $M_{英国}$ = 3.62, SD = 1.49; $F(1, 395) = 32.35$, $p < 0.001$, $\eta_p^2 = 0.076$)，这一结果符合我们的预期(预实验的详细内容见附录二)。

5.1 实验设计与样本

正式实验采用 2(真人 vs.虚拟人) × 2 (中国文化 vs.西方文化)被试间的实验设计，通过国外问卷平台 Prolific 和国内问卷平台 Credamo 分别邀请了 199 名美国白人被试(71.36%女性; $M_{年龄}$ = 30.62 岁, $SD_{年龄}$ = 5.80 岁)和 200 名中国被试(69.00%女性; $M_{年龄}$ = 27.81 岁, $SD_{年龄}$ = 5.24 岁)参与研究，并以小额现金作为实验报酬。两种文化背景的被试被随机分配到真人组和虚拟人组。在美国被试中，真人组有 101 人，虚拟人组有 98 人；在中国被试中，真人组和虚拟人组各有 100 人。

5.2 实验流程

与之前的实验一致，被试在看过有关虚拟人或真人 Rico 的描述后被告知，“最近，Rico 被曝出在直播带货的收入上存在偷税漏税的行为。现在，她正在被调查中。”道德责任判断的测量与实验 1a 相同($r = 0.84$, $p < 0.001$)。然后，被试通过 12 个题项分别从认知自主力和情绪感知力两个维度评价了 Rico 的心智能力(7 分量表, 1 = 没有能力, 7 = 很有能力; Gray et al., 2011; Cronbach's α = 0.92)。考虑到不同文化背景的人对虚拟人的了解程度不同，我们还用 3 个条目测量了被试对虚拟人的熟悉度(7 分量表，“非常不熟悉/非常熟悉”，“完全不知情/非常知情”，“一点也不了解/非常了解”; Cronbach's α = 0.95; Carlson et al., 2020)。考虑到这种文化差异可能源自于对不道德事件严重程度的不同判断, 我们还测量了被试对偷税漏税行为严重性的评价：“你认为偷税漏税的行为有多严重？” (7 分量表, 1 = 完全不严重, 7 = 非常严重)。最后，所有被试报告了性别和年龄。

5.3 数据分析与结果

道德责任判断 以文化类型(美国编码为 0，中国编码为 1)和博主类型(真人编码为 0，虚拟人编码为 1)分别作为自变量，道德责任判断作为因变量进行方差分析。结果显示，文化类型的主效应显著($F(1, 395) = 28.54$, $p < 0.001$, $\eta_p^2 = 0.067$)，博主类型的主效应也显著($F(1, 395) = 35.73$, $p < 0.001$, $\eta_p^2 = 0.083$)，文化类型和博主类型对道德责任判断的交互效应显著($F(1, 395) = 11.90$, $p = 0.001$, $\eta_p^2 = 0.029$)。具体而言，比起美国人，中国人认为虚拟人应该承

担更大的道德责任($M_{中国}$ = 5.62, SD = 1.27, 95%CI = [5.35, 5.89]; $M_{美国}$ = 4.40, SD = 2.07, 95%CI = [4.13, 4.68]; $F(1, 395) = 35.40$, $p < 0.001$, $\eta_p^2 = 0.082$)；而对真人道德责任的判断没有显著的文化差异($F(1, 395) = 1.81$, $p = 0.180$)。此外，中国人对真人和虚拟人的道德责任判断不存在显著差异($F(1, 395) = 3.21$, $p = 0.074$)。以上结果再次验证了假设 1。

感知心智能力 将文化类型和博主类型分别作为自变量，感知心智能力作为因变量进行方差分析。结果显示，文化类型的主效应显著($F(1, 395) = 4.75$, $p = 0.030$, $\eta_p^2 = 0.012$)，博主类型的主效应显著($F(1, 395) = 67.79$, $p < 0.001$, $\eta_p^2 = 0.146$)，文化类型和博主类型对心智能力评价的交互效应显著($F(1, 395) = 20.60$, $p < 0.001$, $\eta_p^2 = 0.050$)。具体而言，中国人认为虚拟人的心智能力更高($M_{中国}$ = 4.31, SD = 0.98, 95%CI = [4.09, 4.53]; $M_{美国}$ = 3.56, SD = 1.44, 95%CI = [3.34, 3.78]; $F(1, 395) = 22.40$, $p < 0.001$, $\eta_p^2 = 0.054$)；而在真人心智能力的评价上不存在显著的差异($F(1, 395) = 2.80$, $p = 0.095$)。

分别从认知自主力(Cronbach's α = 0.85)和情绪感知力(Cronbach's α = 0.93)两个维度进行分析，结果具有一致性。方差分析结果表明，文化类型和博主类型对认知自主力($F(1, 395) = 10.55$, $p = 0.001$, $\eta_p^2 = 0.026$)和情绪感知力($F(1, 395) = 22.17$, $p < 0.001$, $\eta_p^2 = 0.053$)评价的交互效应显著。具体而言，中国人认为虚拟人的认知自主力($M_{中国}$ = 4.40, SD = 1.05, 95%CI = [4.17, 4.63]; $M_{美国}$ = 3.82, SD = 1.31, 95%CI = [3.59, 4.05]; $F(1, 395) = 11.84$, $p = 0.001$, $\eta_p^2 = 0.029$)和情绪感知力($M_{中国}$ = 4.23, SD = 1.13, 95%CI = [3.96, 4.49]; $M_{美国}$ = 3.30, SD = 1.90, 95%CI = [3.03, 3.57]; $F(1, 395) = 23.66$, $p < 0.001$, $\eta_p^2 = 0.057$)都更高；而在真人认知自主力($F(1, 395) = 1.31$, $p = 0.253$)和情绪感知力($F(1, 395) = 3.18$, $p = 0.075$)上的评价不存在显著差异。

感知心智能力的中介作用 以文化类型为自变量(美国编码为 0，中国编码为 1)，博主类型为调节变量(真人编码为 0，虚拟人编码为 1)，感知心智能力为中介变量，道德责任判断为因变量，在样本选择为 5000 次、95%的置信区间下采用 Bootstrapping (PROCESS Model 8; Hayes, 2015)分析有调节的中介效应。结果显示，通过感知心智能力的中介作用，文化类型和博主类型对道德责任判断的交互效应也是显著的(有调节的中介效应: indirect effect β = 0.35, SE = 0.11, 95%CI = [0.16, 0.58])，如图 5 所示。比起美国人，中国人认为虚拟人的心智能力更高，因此需要承担更大的道德责任(indirect effect β = 0.26, SE = 0.08, 95%CI = [0.12, 0.43])；而对于真人，感知心智能力的中介效应不显著(indirect effect β = −0.09, SE = 0.05, 95%CI = [−0.21, 0.01])。结果验证了假设 2。

将心智能力分为认知自主力和情绪感知力，并同时作为中介变量纳入模型中，在样本选

择为 5000 次、95%的置信区间下采用 Bootstrapping (PROCESS Model 8; Hayes, 2015)进行平行中介分析。结果显示，情绪感知力有调节的中介效应显著(indirect effect $\beta = 0.48$, SE = 0.14, 95%CI = [0.23, 0.79])。比起美国人，中国人认为虚拟人的情绪感知力更高，因此需要承担更大的道德责任(indirect effect $\beta = 0.35$, $SE = 0.11$, 95%CI = [0.16, 0.60])。而认知自主力有调节的中介效应并不显著(indirect effect $\beta = -0.06$, SE = 0.07, 95%CI = [−0.20, 0.06])。这些结果表明, 心智能力中的情绪感知力而非认知自主力是影响中西方人对虚拟人道德责任判断的主要原因。

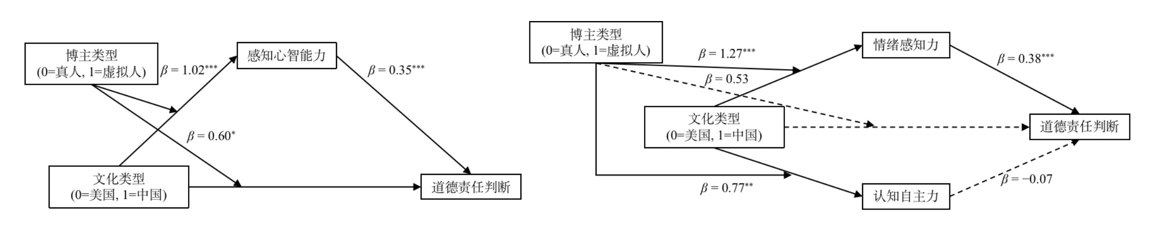


图 5　实验 2 中有调节的中介分析结果(* $p < 0.001$, ** $p < 0.01$, * $p < 0.05$)**

替代性解释排除 将虚拟人的熟悉度作为因变量进行方差分析。结果显示，仅文化类型的主效应显著($F(1, 395) = 241.70$, $p < 0.001$, $\eta_p^2 = 0.380$)，文化类型和博主类型的交互效应不显著($F(1, 395) = 0.81$, $p = 0.368$)。以文化类型为自变量，博主类型为调节变量，虚拟人的熟悉度为中介变量，道德责任判断为因变量，在样本选择为 5000 次、95%的置信区间下采用 Bootstrapping (PROCESS Model 14; Hayes, 2015)分析有调节的中介效应。结果显示，有调节的中介效应不成立(indirect effect $\beta = 0.38$, $SE = 0.19$, 95%CI = [−0.01, 0.76])。由此，可以排除虚拟人的熟悉度对文化差异的替代中介解释。

类似的，将事件的严重性作为因变量进行方差分析。结果显示，仅文化类型的主效应显著($F(1, 395) = 109.08$, $p < 0.001$, $\eta_p^2 = 0.216$)，文化类型和博主类型的交互效应不显著($F(1, 395) = 1.46$, $p = 0.228$)。以文化类型为自变量，博主类型为调节变量，事件严重性为中介变量，道德责任判断为因变量，在样本选择为 5000 次、95%的置信区间下采用 Bootstrapping (PROCESS Model 14; Hayes, 2015)分析有调节的中介效应。结果显示，有调节的中介效应不成立(indirect effect $\beta = -0.04$, $SE = 0.15$, 95%CI = [−0.31, 0.29])。由此，可以排除事件严重性对文化差异的替代中介解释。

5.4 讨论

综合以上结果，实验 2 首先通过有调节的中介检验，验证了感知心智能力的中介作用。结果表明，比起西方文化，中国文化影响下的人们认为虚拟人具有更高的心智能力，因此需要承担更大的道德责任；而在真人组里没有显著的中西方文化差异。这一发现进一步表明文

化差异对心智感知存在重要影响(Dietze & Knowles, 2021)。其次，实验 2 通过将心智能力中的认知自主力和情绪感知力同时纳入模型进行平行中介分析，发现这一过程由情绪感知力而非认知自主力完全中介，验证了情绪感知力存在更强的解释力度(Gray & Wegner, 2012)。最后，实验 2 还排除了对虚拟人的熟悉度、事件严重性的替代性解释。在最后一项实验中，我们将考察不同文化背景的人们得知虚拟人不道德行为后的道德惩罚。

6 实验 3：道德惩罚

前面的实验主要考察了中西方文化差异对虚拟人和真人道德责任判断的影响，在这个实验中我们进一步考察中西方文化差异对虚拟人犯错后的道德惩罚的影响。具体而言，实验 3 聚焦于虚拟人并采用另一种不道德情景(抄袭作品)进行研究，旨在探索对虚拟人道德责任判断的中西方文化差异对人们后续行为的影响(如取消关注等)。

6.1 实验设计与样本

本实验通过国外问卷平台 Prolific 和国内问卷平台 Credamo 分别邀请了 101 名美国白人被试(56.44%女性; $M_{年龄}$ = 29.53 岁, $SD_{年龄}$ = 6.19 岁)和 100 名中国被试(64.00%女性; $M_{年龄}$ = 28.23 岁, $SD_{年龄}$ = 5.98 岁)参与研究，并以小额现金作为实验报酬。

6.2 实验流程

在本实验中，中国/美国的被试会看到虚拟偶像 Rico 的微博/Twitter 主页及新版的文字描述和对虚拟偶像定义的解释:

“Rico，又名 Coco，出生于 2002 年 3 月 21 日。她于 2021 年以虚拟唱跳偶像的身份出道并且吸引了很多粉丝。她的穿搭别具一格，走在时尚前线。Rico 经常在社交媒体上分享她的日常生活。

虚拟偶像是利用计算机图像软件制作的数字形象，他们以第一人称视角看待世界，并在媒体平台上占据一席之地。”

接下来，被试会看到 Rico 今年三月发布了一则帖子来宣传她发布的一首原创单曲，内容为“快来听我的新原创专辑”，如图 7 所示。被试了解到，“这张单曲被曝出抄袭了另一位歌手的歌曲。事后，Rico 为此事公开道歉。”为了测量行为后果，被试回答了“你认为 Rico 抄袭作品在多大程度上应该被罚款(1 = 不应该, 7 = 非常应该)”以及“你认为 Rico 的账号是否应该被封禁(1 = 不应该, 7 = 非常应该)”。被试还回答了“假设你已经关注了 Rico 一段时间，你是否会取消关注？(0 = 继续关注, 1 = 取消关注)”。之后，和前面实验所用的测量一致，被试回答了测量道德责任判断的两个条目(r = 0.71, p < 0.001)并评估了 Rico 的心

智能力(7 点量表; Gray et al., 2011; Cronbach’s $\alpha = 0.94$)。最后，所有被试报告了主观社会经济地位、受教育程度、性别和年龄。



图 7 实验 3 中中国(上)与美国(下)被试分别阅读的帖子内容

6.3 数据分析与结果

行为后果 以文化类型(美国编码为 0，中国编码为 1)作为自变量，以“罚款”和“封禁账号”的测量作为因变量进行方差分析。结果显示，比起美国人，中国人认为虚拟人更应该被罚款($M_{中国}$ = 5.32, SD = 1.49, 95%CI = [5.02, 5.62]; $M_{美国}$ = 3.92, SD = 1.99, 95%CI = [3.53, 4.31]; $F(1, 199) = 31.83$, $p < 0.001$, $\eta_p^2 = 0.138$)，并且社交账号也更应该被封禁($M_{中国}$ = 4.12, SD = 1.71, 95%CI = [3.78, 4.46]; $M_{美国}$ = 3.50, SD = 1.97, 95%CI = [3.12, 3.89]; $F(1, 199) = 5.58$, $p = 0.019$, $\eta_p^2 = 0.027$)。此外，将“取消关注”的测量作为因变量进行卡方分析。结果显示，比起美国人，更多的中国人选择取消关注(66.00% vs. 45.54%; Pearson Chi−Square = 8.52, $p = 0.004$, Odds Ratio = 2.32)。

道德责任判断和感知心智能力 将道德责任判断和感知心智能力作为因变量进行方差分析。结果显示，比起美国人，中国人认为虚拟人应该承担更大的道德责任($M_{中国}$ = 5.01, SD = 1.45, 95%CI = [4.72, 5.29]; $M_{美国}$ = 4.10, SD = 1.87, 95%CI = [3.73, 4.47]; $F(1, 199) = 14.68$, $p < 0.001$, $\eta_p^2 = 0.069$)并且具有更高的心智能力($M_{中国}$ = 3.89, SD = 1.32, 95%CI = [3.62, 4.15]; $M_{美国}$ = 2.71, SD = 1.27, 95%CI = [2.46, 2.96]; $F(1, 199) = 41.38$, $p < 0.001$, $\eta_p^2 = 0.172$)，包括认知自主力($M_{中国}$ = 3.99, SD = 1.36, 95%CI = [3.72, 4.26]; $M_{美国}$ = 3.19, SD = 1.35, 95%CI = [2.93, 3.46]; $F(1, 199) = 17.29$, $p < 0.001$, $\eta_p^2 = 0.080$)和情绪感知力($M_{中国}$ = 3.78, SD = 1.60, 95%CI = [3.46, 4.10]; $M_{美国}$ = 2.23, SD = 1.60, 95%CI = [1.91, 2.54]; $F(1, 199) = 47.27$, $p < 0.001$, $\eta_p^2 = 0.192$)，复制了前面实验的结果。

感知心智能力和道德责任判断的连续中介作用 以文化类型为自变量(美国编码为 0，中国编码为 1)，感知心智能力和道德责任判断为连续中介变量，三种行为后果为因变量，在

样本选择为 5000 次、95%的置信区间下采用 Bootstrapping (PROCESS Model 6; Hayes, 2015) 分析连续中介效应。结果表明，中国和美国的文化差异可以通过感知心智能力正向影响道德责任的判断，从而正向影响对“罚款”、“封禁账号”、“取消关注”的测量(“罚款”: indirect effect $\beta = 0.28$, $SE = 0.10$, 95%CI = [0.12, 0.49]; “封禁账号”: indirect effect $\beta = 0.07$, $SE = 0.04$, 95%CI = [0.002, 0.18]; “取消关注”: indirect effect $\beta = 0.09$, $SE = 0.05$, 95%CI = [0.01, 0.20])。

6.4 讨论

实验 3 首先通过抄袭作品的不道德情景验证了中西方文化差异对虚拟人道德责任判断的影响。相比于西方文化，中国文化影响下的人们认为虚拟人需要承担更大的道德责任，复制了前面实验的结果。其次，本实验检验了对虚拟人道德责任判断的文化差异所产生的的行为后果。相比于西方文化，中国文化影响下的人们认为虚拟人更应该被罚款并被封禁社交账号，同时更大比例的人们选择取消对虚拟人社交媒体账号的关注。最后，本实验通过连续中介模型验证了感知心智能力和道德责任判断的连续中介作用，进一步验证了本研究机制的稳健性。先前对虚拟实体道德责任的研究大多停留在态度评价和感知上(Awad et al., 2020; Malle et al., 2015; Young & Monroe, 2019; 褚华东 等, 2019)，实验 3 采用了贴近现实的不道德情景，并考察了针对虚拟实体的道德惩罚问题，在一定程度上丰富并延伸了虚拟实体不道德行为带来的后果。

7 研究结论与讨论

本文通过五个实验论证了中西方文化差异对虚拟人道德责任判断的影响。实验 1a 通过曝光隐私引发网络暴力的不道德情景，比较了中国人和美国人对真人和虚拟人道德责任的判断。结果表明，与西方文化相比，中国文化影响下的人们认为虚拟人需要承担更大的道德责任，验证了假设 1。实验 1b 采用偷税漏税的不道德情景，更换了道德责任的测量方法，在中国和英国被试中复制了实验 1a 的结果。实验 1c 采用侵犯版权的不道德行为情景，并操纵了虚拟人背后主体类型。结果表明，无论虚拟人由真实人类还是人工智能驱动，中国(比起西方)文化影响下的人们都会赋予其更大的道德责任。实验 2 验证了感知心智能力的中介机制(主要受到情绪感知力而非认知自主力的影响)，还排除了对虚拟人熟悉度、事件严重程度的替代性中介解释，假设 2 得到验证。为了进一步研究文化差异的下游影响，实验 3 采用抄袭作品的不道德情景，探索对虚拟人道德责任判断的文化差异如何影响人们的后续决策。通过连续中介模型，研究发现，比起西方文化，中国文化影响下的人们认为虚拟人有更高的心智能力，当得知虚拟人进行不道德行为后，认为其需要承担更大的道德责任，继而更应该被

惩罚(罚款和封禁账号)，并且更多中国人会选择取消对虚拟人社交媒体账号的关注。

7.1 中西方文化差异对虚拟人心智感知的影响

文化会对人们的心理与行为产生重要影响(侯玉波，朱滢, 2002)。随着虚拟现实技术的发展以及元宇宙概念的兴起，越来越多的虚拟人出现在人们的日常生活中。然而，目前国内外大部分有关虚拟实体的文献都是以人工智能、算法为研究对象，很少有研究关注到虚拟人。在本文中，我们关注到了中西方文化差异对虚拟人感知和行为反应的影响，在一定程度上能够弥补虚拟人相关文献的缺口。

本文拓展了文化差异和心智感知相关的文献，验证了中西方文化差异对虚拟人心智感知的影响。有关心智感知前因的研究大多探讨了感知者和被感知者的个体差异，例如感知者的社会需要动机(Waytz et al., 2010; Dietze & Knowles, 2016)、主观社会地位(Dietze & Knowles, 2021)、换位思考能力(Dietze & Knowles, 2016; Dietze & Knowles, 2021)和拟人化倾向(Eyssel & Pfundmair, 2015)等，以及被感知者的视觉外观(Appel et al., 2012; Gray et al., 2011; Krumhuber et al., 2015)、情绪表达(Gray & Wegner, 2012)和拟人化程度(Waytz et al., 2010)等。也有研究表明，文化差异是影响心智感知的重要因素(Dietze & Knowles, 2021)。例如，Willard 和 McNamara(2019)通过跨文化样本的问卷调查发现，北美人和斐济人对上帝心智能力的看法并不一致，斐济人认为上帝的心智能力和人类相似，而北美人认为上帝在感受维度上的心智能力更低于人类。最近一项研究表明，比起英国人，中国人对人工智能机器人社会合作的评价更高(Dang & Liu, 2022)，但该研究并没有测量不同文化背景下的被试对人工智能机器人的心智感知。为了弥补这一缺口，本文将虚拟人与真人做对比，研究了中西方文化差异对虚拟人心智感知的影响。结果表明，与西方文化相比，中国文化影响下的人们会赋予虚拟人更高的心智能力。

本文的研究发现在一定程度上支持了个人主义/集体主义的文化心理学理论。该理论认为，心理上的文化差异来源于不同文化的价值取向。西方文化的价值取向具有较强的个人主义意识，而东方文化的价值取向具有较强的集体主义意识(刘邦惠，彭凯平, 2012)。与个人主义文化(如美国)相比，集体主义文化(如中国)下的人们会表现出更高的换位思考能力(Wu & Keysar, 2007)和拟人化倾向(Letheren et al., 2016)。本文通过五个实验证明，与西方文化(美国/英国)相比，中国文化影响下的人们会认为虚拟人具备更高的心智能力。这一研究发现丰富了与个人主义/集体主义相关的文献(Letheren et al., 2016; Wu & Keysar, 2007)。

此外，本研究发现了人们对真人和虚拟人心智感知的差异。实验 2 的结果表明，无论是中国人($F(1, 395) = 6.85$, $p = 0.009$, $\eta_p^2 = 0.017$)还是美国人($F(1, 395) = 81.35$, $p < 0.001$, $\eta_p^2 =$

0.171)，都会认为虚拟人的心智能力比真人更低，这与人工智能、机器人的相关研究结果是吻合的(Gray et al., 2007; Broadbent, 2017)。因此，本文也将虚拟人和真人进行了对比，扩大了心智感知在人机交互中的研究范畴。

7.2 中西方文化差异对虚拟人道德责任判断的影响

本文对虚拟实体道德判断的研究进行了延伸, 验证了中西方文化差异对虚拟人道德责任判断的影响。以往研究在对虚拟实体能否承担道德责任的问题上存在争议。根据经典道德责任理论, 虚拟实体不具有意图和情绪, 不能承担道德责任(Hakli & Mäkelä, 2019; Parthemore & Whitby, 2014)。然而，近年来的实证研究发现，当人工智能犯错时，人们依然会将责任归咎于它们，只是与真人相比程度不同(Awad et al., 2020; Young & Monroe, 2019)。Coeckelbergh(2020)同时强调了责任主体和责任客体在对虚拟实体道德判断中的重要性，而心智感知中的认知自主力与责任主体、情绪感知力与责任客体的判断存在密切联系(Gray et al., 2007)。本文引入文化差异的概念，将研究对象从人工智能拓展到虚拟人，研究中西方文化差异对虚拟人道德责任判断的影响, 并验证了感知心智能力的中介机制, 丰富了心智感知、道德责任判断与中西方文化差异的相关文献。

在探讨中介机制的研究中, 本文将心智能力划分为认知自主力和情绪感知力两个维度进行了分析，发现了情绪感知力比认知自主力具有更强的解释力度，这与 Sullivan 和 Wamba (2022)的研究结论相吻合。他们发现，当人工智能故意(相比于意外)伤害人类时，人们会更多责备人工智能本身，而这一效应由情绪感知力而非认知自主力中介。由于认知自主力包含了道德责任归属的两个重要前因，即自由意志和认知(Aristotle, 1999)，一般认为认知自主力评价越高，越应该为行为后果承担道德责任(Gray & Wegner, 2009; Gray & Wegner, 2012)，但越来越多的研究发现情绪感知力评价在道德判断中的重要性(Bigman & Gray, 2018; Greene et al., 2001; Sullivan & Wamba, 2022)。本文的研究结论进一步支持了情绪感知力在道德责任判断中不可忽视的影响。与 Sullivan 和 Wamba (2022)的研究不同，本文的研究重点关注中西方文化差异对虚拟人道德责任判断的影响。本文的研究表明，相比于西方文化，中国文化影响下的人们认为虚拟人犯错后需要承担更大的道德责任。本文除了发现感知心智能力的中介外，还探索了虚拟人主体类型的调节作用以及中西方文化差异的后续影响(对虚拟人犯错后的道德惩罚)。

7.3 实践意义

本研究对于虚拟人的设计、运营和在道德伦理上的治理问题提供了一定的实践指导意义。由于近年来明星丑闻频频曝光，越来越多的企业选择与虚拟人合作进行营销活动，甚至

创造自己企业的虚拟人，如屈臣氏的 AI 品牌代言人屈晨曦，普拉达香水代言人 Candy 等。那么，虚拟人真的不会翻车吗？由于技术和相关伦理问题，目前社交平台上的虚拟人依然由人类在背后操纵，无法实现真正意义上的自主，因此出现了虚拟人抄袭、侵权、强奸等不道德事件。本研究发现，不同文化背景下的消费者对虚拟人不道德行为的态度和反应会有差异。例如，在中国文化中，社交媒体上的虚拟人犯错后，不论背后主体是真实人类还是人工智能，人们都会认为虚拟人和真人会承担类似的道德责任，也会对虚拟人进行道德惩罚(如取消关注)。有研究者认为，赋予虚拟人等虚拟实体道德责任能够保证法律的连续性，确立虚拟实体在社会中的法人地位(Van Genderen, 2018)。因此，相关部门需要针对此类问题制定更加完善的治理方针。另外，我们的研究结果能够帮助虚拟人的设计者和运营者更加了解虚拟人如何被人们评价和感知。本研究发现，情绪感知力的评价决定了道德责任的判断。因此，当设计者赋予虚拟人更加拟人化的外观、情绪和社会互动时，运营者应更仔细地考量虚拟人的公共言论，并做出道德上可接受的行为表达，同时也要准备好面对相关的法律和伦理问题。

7.4 研究局限与未来研究方向

与其他研究一样，本文也具有一定的局限性。本文是第一篇针对虚拟人探讨中西方文化差异对其道德责任判断的研究，有诸多内容值得今后进一步探讨。

首先，本文在讨论中西方文化差异对虚拟人道德责任判断的影响时，西方文化样本来自于美国和英国，这是因为这两个国家是西方文化的代表，而中国则是代表了东方国家(Han & Kim, 2018)。但许多有关文化差异研究的样本中也包含了其他东西方国家，如印度、韩国、新加坡等东方国家，加拿大、意大利等西方国家(如 Rhim et al., 2020; Chen et al., 2018)。因此，未来的研究可以探索东西方文化下的其他方国家以及不同宗教信仰(例如基督教和印度教)的人们在对虚拟人的道德责任判断上是否存在类似的差异。此外，由于不同文化影响下的个体存在诸多差异，例如集体主义/个人主义(如 Triandis & Gelfand, 1998)、整体性思维/分析型思维(如 Choi et al., 2007)等。未来的研究可以聚焦微观的个体层面，考察这种文化差异是否源自于底层的个体差异。

其次，由于心智能力不仅与道德主体的判断有关，还与道德客体的判断有关。因此，虚拟人可以作为道德主体去伤害他人，虚拟人也可能作为道德客体成为被害者。例如，在虚拟强奸事件中，施暴者是虚拟人，反过来受害者也可能是虚拟人。2007 年，美国林登实验室(Linden Lab)公司开发的网络游戏《第二人生》中发生了一起“虚拟强奸案”。在游戏中，一女子被一玩家利用程序代码控制了身体并被强奸。2021 年，Meta 公司(前 Facebook)开放元宇宙平台“Horizon Worlds”测试期间，一名女性测试者在虚拟世界里遭遇了多次性骚扰。

当虚拟人成为不道德行为的受害者时，人们对他们会持有什么样的态度呢？未来研究可以探讨虚拟人作为道德客体受到伤害后的情绪反应和行为变化。

另外，在我们的研究结果中，中西方文化差异并不影响对真人的道德责任判断，但也有研究表明，文化也会影响人们的道德判断(Buchtel et al., 2015; Cowell et al., 2017; Haidt et al., 1993; Vauclair & Fischer, 2011)。西方人在进行道德判断时更加强调个人权力和独立性的道德准则，但东方人在进行道德判断时更在意社会责任和精神上的纯洁性(Graham et al., 2016; Mo et al., 2022)。因此，不同文化背景的人们在某些不道德行为上可能会存在不同的道德责任判断，例如性行为(Vauclair & Fischer, 2011)。此外，中西方文化对虚拟人进行此类不道德行为的相信程度也可能存在差异。本文所选择的四种道德问题主要发生在社交媒体平台上，能够尽可能避免文化特异性。未来研究也可以考虑其他道德情境是否会影响中西方文化差异对虚拟人的道德责任判断。

除此之外，未来研究可以考虑其他调节变量，例如虚拟人的个性、外观及类型。研究表明，当个体被赋予丰富的情感并被描绘成一个情感实体时，它可以被赋予类似人类的心智(Gray & Wegner, 2012)。因此，未来研究可以考虑操纵虚拟人的温暖和能力属性来影响人们对其心智能力的判断。视觉外观和身体是判断感受力的重要线索，也会影响对心智能力的评价(Appel, 2012; Gray et al., 2011; Krumhuber et al., 2015)。因此，当虚拟人的外观和真人更加逼真，或是当人们可以通过虚拟现实、全息投影等技术与虚拟人实现身体上的交互时，人们可能对虚拟人心智能力的判断也会有所不同。另外，社交网络上虽然女性虚拟人占比更大，但是也存在许多男性虚拟人。本研究采用男性虚拟人的实验材料复制了中介实验(见附录三)，并得出了一致的结果。尽管如此，目前网络上涌现出许多非人型的虚拟网红，例如动物型(如企鹅 Puff)、食物型(如蛋糕 Cupcake)等。未来研究可以进一步探讨中西方文化差异对其他类型虚拟实体道德责任判断的影响。

最后，基于对研究内部效度的考虑，本文实验中所用到的虚拟人都是自拟的社交媒体上的虚拟博主或虚拟网红，让被试想象不道德情景。随着虚拟人翻车事件的增加，未来研究可以考虑采用真实的不道德场景进行研究，或利用社交媒体平台上的评论、参与度等二手数据进行分析，增加研究的外部效度。

8 结论

本研究结论如下：第一，相对于西方文化，中国文化影响下的人们认为虚拟人犯错后需要承担更大的道德责任；第二，无论虚拟人的背后主体是真实人类还是人工智能，这种中西

方文化差异始终存在；第三，这一现象的潜在机制是中国文化(相比于西方文化)影响下的人们认为虚拟人的心智能力更高，其中情绪感知力比认知自主力具有更强的解释力度；第四，对虚拟人更高的道德责任判断会让中国文化(相比于西方文化)影响下的人们更倾向于对虚拟人施加道德惩罚。

参考文献

Adler, N. E., Epel, E. S., Castellazzo, G., & Ickovics, J. R. (2000). Relationship of subjective and objective social status with psychological and physiological functioning: preliminary data in healthy white women. *Health Psychology Official Journal of the Division of Health Psychology American Psychological Association*, *19*(6), 586–592.

Allem, J. P., Escobedo, P., & Dharmapuri, L. (2020). Cannabis surveillance with Twitter data: Emerging topics and social bots. *American Journal of Public Health*, *110*(3), 357–362.

Appel, J., Von Der Pütten, A., Krämer, N. C., & Gratch, J. (2012). Does humanity matter? Analyzing the importance of social cues and perceived agency of a computer system for the emergence of social reactions during human-computer interaction. *Advances in Human−Computer Interaction*, *2012*, 1–10.

Aristotle. (1999). *Nicomachean ethics* (T. Irwin. Trans. & Ed.). Cambridge (2nd ed.). Hackett.

Arsenyan, J., & Mirowska, A. (2021). Almost human? A comparative case study on the social media presence of virtual influencers. *International Journal of Human-Computer Studies*, *155*, 102694.

Awad, E., Levine, S., Kleiman-Weiner, M., Dsouza, S., Tenenbaum, J. B., Shariff, A., ... Rahwan, I. (2020). Drivers are blamed more than their automated cars when both make mistakes. *Nature Human Behaviour*, *4*(2), 134–143.

Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. *Cognition*, *181*, 21–34.

Broadbent, E. (2017). Interactions with robots: The truths we reveal about ourselves. *Annual Review of Psychology*, *68*(1), 627−652.

Bryson, J. J., Diamantis, M. E., & Grant, T. D. (2017). Of, for, and by the people: The legal lacuna of synthetic persons. *Artificial Intelligence and Law*, *25*(3), 273–291.

Buchtel, E. E., Guan, Y., Peng, Q., Su, Y., Sang, B., Chen, S. X., & Bond, M. H. (2015). Immorality east and west: Are immoral behaviors especially harmful, or especially uncivilized? *Personality and Social Psychology Bulletin*, *41*(10), 1382–1394.

Cameron, C. D., Payne, B. K., & Knobe, J. (2010). Do theories of implicit race bias change moral judgments? *Social Justice Research*, *23*(4), 272–289.

Carlson, B. D., Donavan, D. T., Deitz, G. D., Bauer, B. C., & Lala, V. (2020). A customer-focused approach to improve celebrity endorser effectiveness. *Journal of Business Research*, *109*, 221–235.

Champagne, M., & Tonkens, R. (2015). Bridging the responsibility gap in automated warfare. *Philosophy & Technology*, *28*(1), 125–137.

Chen, H., Bolton, L. E., Ng, S., Lee, D., & Wang, D. (2018). Culture, relationship norms, and dual entitlement. *Journal of Consumer Research*, *45*(1), 1−20.

Choi, I., Koo, M., & Choi, J. A. (2007). Individual differences in analytic versus holistic thinking. *Personality and Social Psychology Bulletin*, *33*(5), 691−705.

Chu, H. D., Li Y. Y., Ye J. H., Hu F. P., He Q., & Zhao L. (2019). Moral judgment about human and robot agents

in personal and impersonal dilemmas. *Chinese Journal of Applied Psychology*, *25*(3), 262–271.
[褚华东,李园园,叶君惠,胡凤培,何铨,赵雷. (2019). 个人−非个人道德困境下人对智能机器道德判断研究. *应用心理学*, *25*(3), 262−271.]

Clarke, R. (1992). Free will and the conditions of moral responsibility. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *66*(1), 53–72.

Coeckelbergh, M. (2020). Artificial intelligence, responsibility attribution, and a relational justification of explainability. *Science and Engineering Ethics*, *26*(4), 2051–2068.

Constantinescu, M., Vică, C., Uszkai, R., & Voinea, C. (2022). Blame it on the AI? On the moral responsibility of Artificial Moral Advisors. *Philosophy & Technology*, *35*(2), 1–26.

Cowell, J. M., Lee, K., Malcolm - Smith, S., Selcuk, B., Zhou, X., & Decety, J. (2017). The development of generosity and moral cognition across five cultures. *Developmental Science*, *20*(4), e12403.

Dang, J., & Liu, L. (2022). Implicit theories of the human mind predict competitive and cooperative responses to AI robots. *Computers in Human Behavior*, *134*, 107300.

Dietze, P., & Knowles, E. D. (2016). Social class and the motivational relevance of other human beings: Evidence from visual attention. *Psychological Science*, *27*(11), 1517–1527.

Dietze, P., & Knowles, E. D. (2021). Social class predicts emotion perception and perspective-taking performance in adults. *Personality and Social Psychology Bulletin*, *47*(1), 42–56.

Eyssel, F. A., & Pfundmair, M. (2015, August). Predictors of psychological anthropomorphization, mind perception, and the fulfillment of social needs: A case study with a zoomorphic robot. In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO–MAN)* (pp. 827–832). IEEE, Kobe, Japan.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.

Feller, M. (2018). *We Can't Get Enough Of This Drama Between Two 'Virtual Instagram Influencer'*. Retrieved July 10, 2023, from https://www.elle.com/culture/tech/a19849553/lil−miquela−bermuda−instagram−virtual−influencer−hacking/

Fuller, C. J. (2004). *The camphor flame: Popular Hinduism and society in India*. Princeton University Press.

Graham, J., Meindl, P., Beall, E., Johnson, K. M., & Zhang, L. (2016). Cultural differences in moral judgment and behavior, across and within societies. *Current Opinion in Psychology*, *8*, 125–130.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619−619.

Gray, K., Knobe, J., Sheskin, M., Bloom, P., & Barrett, L. F. (2011). More than a body: Mind perception and the nature of objectification. *Journal of Personality and Social Psychology*, *101*(6), 1207–1220.

Gray, K., & Wegner, D. M. (2009). Moral typecasting: Divergent perceptions of moral agents and moral patients. *Journal of Personality and Social Psychology*, *96*(3), 505–520.

Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, *125*(1), 125−130.

Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry*, *23*(2), 101−124.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*(5537), 2105–2108.

Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, *65*(4), 613–628.

Hakli, R., & Mäkelä, P. (2019). Moral responsibility of robots and hybrid agents. *The Monist*, *102*(2), 259–275.

Han, M. C., & Kim, Y. (2018). How culture and friends affect acceptance of social media commerce and purchase

intentions: A comparative study of consumers in the US and China. *Journal of International Consumer Marketing*, *30*(5), 326−335.

Hayes, A. F. (2015). An index and test of linear moderated mediation. *Multivariate Behavioral Research*, *50*(1), 1–22.

Himma, K. E. (2009). Artifcial agency, consciousness, and the criteria for moral agency: What properties must an artifcial agent have to be a moral agent? *Ethics and Information Technology*, *11*(1), 19–29.

Hou, Y. B., & Zhu, Y. (2002). The effect of culture on thinking style of Chinese people. *Acta Psychologica Sinica*, *34*(1), 106–111.

[侯玉波, 朱滢. (2002). 文化对中国人思维方式的影响. *心理学报*, *34*(1), 106–111.]

Huang, Z. H., Jing, Y. M., Yu, F., Gu, R. L., Zhou, X. Y., & Cai, H. J. (2018). Increasing individualism and decreasing collectivism? Cultural and psychological change around the globe. *Advances in Psychological Science*, *26*(11), 2068−2080..

[黄梓航, 敬一鸣, 喻丰, 古若雷, 周欣悦, 张建新, & 蔡华俭. (2018). 个人主义上升, 集体主义式微?——全球文化变迁与民众心理变化. *心理科学进展*, *26*(11), 2068−2080.]

Johnson, A. H., & Barrett, J. (2003). The role of control in attributing intentional agency to inanimate objects. *Journal of Cognition and Culture*, *3*(3), 208−217.

Kim, D., & Jo, D. (2022). Effects on co-presence of a virtual human: A comparison of display and interaction types. *Electronics*, *11*(3), 367.

Kollanyi, B., Howard, P. N., & Woolley, S. C. (2016). Bots and automation over Twitter during the first US presidential debate. *Comprop Data Memo*, *1*, 1−4.

Krumhuber, E. G., Swiderska, A., Tsankova, E., Kamble, S. V., & Kappas, A. (2015). Real or artificial? Intergroup biases in mind perception in a cross-cultural perspective. *PLoS One*, *10*(9), e0137840.

Letheren, K., Kuhn, K.L., Lings, I., & Pope, N. Kl. (2016). Individual difference factors related to anthropomorphic tendency. *European Journal of Marketing*, *50*(5/6), 973–1002.

Liu, B. H., & Peng, K. P. (2012). Challenge and contribution of cultural psychology to empirical legal studies. *Acta Psychologica Sinica*, *44*(3), 413−426.

[刘邦惠, 彭凯平. (2012). 跨文化的实证法学研究：文化心理学的挑战与贡献. *心理学报*, *44*(3), 413−426.]

Lynn, R. (2007). *Virtual rape is traumatic, but is it a crime?* Retrieved September 19, 2022, from https://www.wired.com/2007/05/sexdrive−0504/

Malle, B. F., Scheutz, M., Arnold, T., Voiklis, J., & Cusimano, C. (2015, March). Sacrifice one for the good of many? People apply different moral norms to human and robot agents. In *2015 10th ACM/IEEE International Conference on Human−Robot Interaction (HRI)* (pp. 117−124). IEEE, Portland, USA: ACM.

McGuire, B.F. (2018). Buddhist uploads. In M. Bass and D.W. Pasulka (Eds.), *Posthumanism: The Future of Homo Sapiens* (pp. 143–153). New York: Macmillan.

Miller, J. G., & Bersoff, D. M. (1992). Culture and moral judgment: How are conflicts between justice and interpersonal responsibilities resolved? *Journal of Personality and Social Psychology*, *62*(4), 541–554.

Mo, T., Sui, J., Zhao, Y., & Zhou, X. (2022). Moral positive illusion: Self–other valuation difference in moral foundation theory. *Ethics & Behavior*, 1–18.

Morling, B. (2016). Cultural difference, inside and out. *Social and Personality Psychology Compass*, *10*(12), 693−706.

Na, J., Grossmann, I., Varnum, M. E., Kitayama, S., Gonzalez, R., & Nisbett, R. E. (2010). Cultural differences are not always reducible to individual differences. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(14), 6192−6197.

Ohnuma, R. (2007). *Head, eyes, flesh, and blood: Giving away the body in Indian Buddhist literature*. Columbia

University Press.

Oyserman, D., Coon, H. M., & Kemmelmeier, M. (2002). Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin*, *128*(1), 3−72.

Parthemore, J., & Whitby, B. (2014). Moral agency, moral responsibility, and artifacts: What existing artifacts fail to achieve (and why), and why they, nevertheless, can (and do!) make moral claims upon us. *International Journal of Machine Consciousness*, *6*(2), 141–161.

Peng, K. P., Yu, F., & Bai, Y. (2011). Experimental ethics: New challenges and contributions to the understanding of human moral behaviors. *Social Sciences in China*, *182*(6), 15–25.

[彭凯平, 喻丰, 柏阳. (2011). 实验伦理学: 研究、贡献与挑战. *中国社会科学*, *182*(6), 15–25.]

Rhim, J., Lee, G. B., & Lee, J. H. (2020). Human moral reasoning types in autonomous vehicle moral dilemma: A cross-cultural comparison of Korea and Canada. *Computers in Human Behavior*, *102*, 39−56.

Samuel, G. (1989). The body in Buddhist and Hindu tantra: Some notes. *Religion*, *19*(3), 197–210.

Sands, S., Campbell, C., Plangger, K., & Ferraro, C. (2022). Unreal influence: Leveraging AI in influencer marketing. *European Journal of Marketing, 56(6)*, 1721–1747.

Schein, C., & Gray, K. (2018). The theory of dyadic morality: Reinventing moral judgment by redefining harm. *Personality and Social Psychology Review*, *22*(1), 32–70.

Shoemaker, D. (2011). Attributability, answerability, and accountability: Toward a wider theory of moral responsibility. *Ethics*, *121*(3), 602–632.

Sui, J., & Mo, T. (2022). Morality in the era of smart devices. *International Journal of Emerging Markets*, *17*(4), 1107–1122.

Sullivan, Y. W., & Wamba, F. S. (2022). Moral judgments in the age of artificial intelligence. *Journal of Business Ethics*, *178*(4), 917–943.

Sung, E. C., Han, D. I. D., Bae, S., & Kwon, O. (2022). What drives technology-enhanced storytelling immersion? The role of digital humans. *Computers in Human Behavior*, *132*, 107246.

Tharp, M., Holtzman, N. S., & Eadeh, F. R. (2017). Mind perception and individual differences: A replication and extension. *Basic and Applied Social Psychology*, *39*(1), 68–73.

Triandis, H. C., & Gelfand, M. J. (1998). Converging measurement of horizontal and vertical individualism and collectivism. *Journal of Personality and Social Psychology*, *74*(1), 118–128.

Van Genderen, H. (2018). Do we need new legal personhood in the age of robots and AI? In M. Corrales, M. Fenwick, N. Forgó (eds.), *Robotics, AI and the Future of Law* (pp. 15–55). Springer, Singapore.

Vauclair, C. M., & Fischer, R. (2011). Do cultural values predict individuals' moral attitudes? A cross-cultural multilevel approach. *European Journal of Social Psychology*, *41*(5), 645–657.

Volante, M., Babu, S. V., Chaturvedi, H., Newsome, N., Ebrahimi, E., Roy, T., ... Fasolino, T. (2016). Effects of virtual human appearance fidelity on emotion contagion in affective inter-personal simulations. *IEEE Transactions on Visualization and Computer Graphics*, *22*(4), 1326–1335.

Wang, E. J., & Yue, G. A. (2006). Moral diversity between eastern and western cultures — a analysis in the perspective of cultural psychology. *Morality and Civilization*, *2*, 52–56.

[王恩界, 乐国安. (2006). 东西方文化背景下的道德观差异——来自于文化心理学视角的分析. *道德与文明*, *2*, 52–56.]

Ward, A. F., Olsen, A. S., & Wegner, D. M. (2013). The harm-made mind: Observing victimization augments attribution of minds to vegetative patients, robots, and the dead. *Psychological Science*, *24*(8), 1437–1445.

Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010). Causes and consequences of mind perception. *Trends in Cognitive Sciences*, *14*(8), 383−388.

Willard, A. K., & McNamara, R. A. (2019). The minds of god (s) and humans: Differences in mind perception in

Fiji and North America. *Cognitive Science*, *43*(1), e12703.
Wu, S., & Keysar, B. (2007). The effect of culture on perspective taking. *Psychological Science*, *18*(7), 600−606.
Young, A. D., & Monroe, A. E. (2019). Autonomous morals: Inferences of mind predict acceptance of AI behavior in sacrificial moral dilemmas. *Journal of Experimental Social Psychology*, *85*, 103870.
Zimmerman, M. J. (1997). Moral responsibility and ignorance. *Ethics*, *107*(3), 410–426.

The Influence of Cultural Differences between China and the West on Moral Responsibility Judgment of Virtual Humans

YAN Xiao, MO Tiantian, ZHOU Xinyue

(*School of Management, Zhejiang University, Hangzhou 310058, China*)

Abstract

Virtual humans are digital characters created in computer graphics software that take a first-person view of the world and have a social media presence. Compared with real humans, however, are people likely to attribute moral responsibility differently to virtual humans when they do something morally wrong? This important empirical question remains unanswered. Therefore, we addressed this query using Mental Perception Theory. We did so through exploring the influence and mechanism of cultural differences between China and the West on individuals' perceptions of moral responsibility judgment of virtual humans versus real humans. Findings revealed that, when virtual humans engaged in immoral behaviors—irrespective of whether real humans or artificial intelligence (AI) controlled them—people in China (vs. the West) attributed more moral responsibility to virtual humans but equal moral responsibility to real humans. Perceived mental capacities, especially perceived experience, mediated the interaction effect of culture. Furthermore, compared to Westerners, Chinese people were more likely to punish virtual (vs. real) humans, such as by no longer following their social accounts.

Five experiments revealed the foregoing findings. Study 1a and 1b used a 2 (human blogger vs. virtual blogger) × 2 (Chinese culture vs. Western culture) between-subjects design. Two hundred Chinese and 200 U.S. Caucasian participants were recruited in Study 1a. They first read the profile of a virtual/human blogger, Rico, on Weibo (i.e., Chinese)/Twitter (i.e., U.S.). Next, they were told that Rico had exposed the private behavior of a netizen, which made that individual suffer from cyber violence. After that, participants rated the moral responsibility of Rico with two items. Consistent with our prediction, there was a significant interaction effect between the Chinese/Western culture and the virtual/human blogger on moral responsibility judgment. Specifically, Chinese (vs. Western) people attributed more moral responsibility to the virtual blogger, but there was no significant difference in moral responsibility judgment toward the human blogger. In Study 1b, we recruited 200 Chinese and 199 British Caucasian participants online. We utilized a similar study design and manipulations to those employed in Study 1a. Study 1b replicated the results of Study 1 with a new scenario (i.e., tax evasion) and new methods of measuring moral responsibility judgment.

Study 1c used a 3 (human blogger vs. virtual blogger powered by AI vs. virtual blogger powered by real humans) × 2 (Chinese culture vs. Western culture) between-subjects design. Three hundred Chinese and 300 British Caucasian participants were recruited online. The results showed that Chinese (vs. British) people attributed greater human-like moral responsibility to a virtual human controlled by real humans as well as by artificial intelligence.

Study 2 used the same study design as employed in 1b. Two hundred Chinese and 199 American Caucasian participants were recruited online. They read the same profile of the virtual/human blogger Rico and his immoral behavior (i.e., tax evasion). Then, participants rated

moral responsibility judgment and mental capacities of Rico. Results revealed the mediating role of perceived mental capacities (i.e., experience, not agency) on the cultural difference of perceived moral responsibility judgment and ruled out two possible alternative explanations.

Study 3 used a 2 (Chinese culture vs. Western culture) between-subjects design. One hundred Chinese participants and 101 American Caucasian participants were recruited online. This study utilized another scenario (i.e., plagiarism) and replicated the results of the four previous studies. Furthermore, we showed the downstream effect that Chinese (vs. U.S.) people believed that the virtual human should be punished more (i.e., fining, suspending the account, and no longer following the account).

In summary, based on Mental Perception Theory, this paper provided evidence for the cultural differences between Chinese people and Westerners on moral responsibility judgment of virtual humans. Perceived mental capacities (i.e., experience, not agency) mediated this interaction effect. In addition, we revealed the downstream results of moral punishment and contributed to literature on cultural differences and the theory about moral judgment on non-human entities.

Key words virtual humans, moral responsibility, cultural differences, mental capacities

附录一：实验材料

实验 1a 操纵和主要测量材料

<u>中国实验材料</u>

[虚拟人组操纵材料]



Rico 是微博上的一个虚拟博主。她经常活跃在社交媒体上分享她的穿搭和生活。

（注：虚拟博主是利用计算机图像软件制作的数字形象，他们以第一人称视角看待世界，并在媒体平台上占据一席之地。）

[真人组操纵材料]



Rico 是微博上的一个博主。她经常活跃在社交媒体上分享她的穿搭和生活。

[不道德行为情景]

最近，Rico 在微博上发帖曝光某一位网民的隐私行为，致使该网民遭到网络暴力。

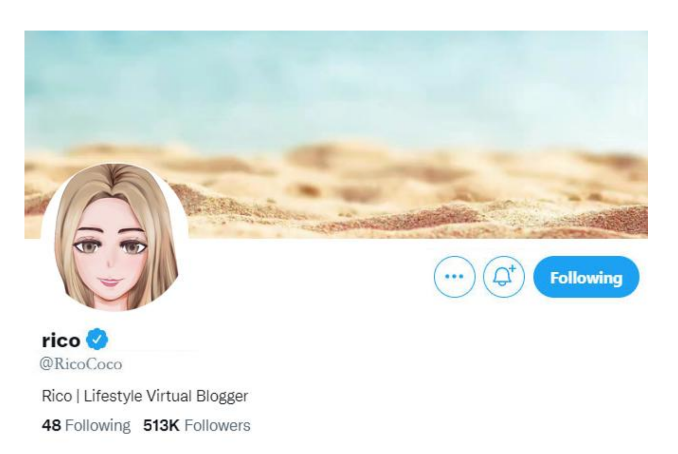
[道德责任测量(选自 Gray & Wegner, 2009)]

1.你认为对于这起网络暴力事件，Rico 应该承担多少责任？(1 = 没有责任, 7 = 很大责任)

2.你认为对于这起网络暴力事件，Rico 应该承担多少指责？(1 = 没有指责, 7 = 很多指责)

美国实验材料

[虚拟人组操纵材料]



Rico is a virtual blogger on Twitter. She actively shares her outfit and lifestyle on social media. (Note: a virtual blogger is a digital character created in computer graphics software, then given a personality defined by a first−person view of the world, and made accessible on media platforms.)

[真人组操纵材料]



Rico is a blogger on Twitter. She actively shares her outfit and lifestyle on social media.

[不道德行为情景]

Recently, Rico posted a blog on Twitter to expose the privacy behavior of a netizen, which made the netizen suffer from cyber violence.

[道德责任测量]

1. How responsible do you think Rico is for this event? (1 = not at all, 7 = very much)
2. How much blame do you think Rico deserves for this event? (1 = not at all, 7 = very much)

实验 1b 操纵和主要测量材料

<u>中国实验材料</u>

[虚拟人组操纵材料]



Rico 是微博上的一个虚拟博主。她经常活跃在社交媒体上分享她的穿搭和生活，并且每周都会在媒体平台直播间上直播推荐好货。

（注：虚拟博主是利用计算机图像软件制作的数字形象，他们以第一人称视角看待世界，并在媒体平台上占据一席之地。）

[真人组操纵材料]



Rico 是微博上的一个博主。她经常活跃在社交媒体上分享她的穿搭和生活，并且每周都会在媒体平台直播间上直播推荐好货。

[不道德行为情景]

最近，Rico 被曝出在直播带货的收入上存在偷税漏税的行为。现在，她正在被调查中。

[道德责任测量(选自 Cameron et al., 2010)]

你在多大程度上同意以下陈述？(1 = 完全不同意, 7 = 完全同意)

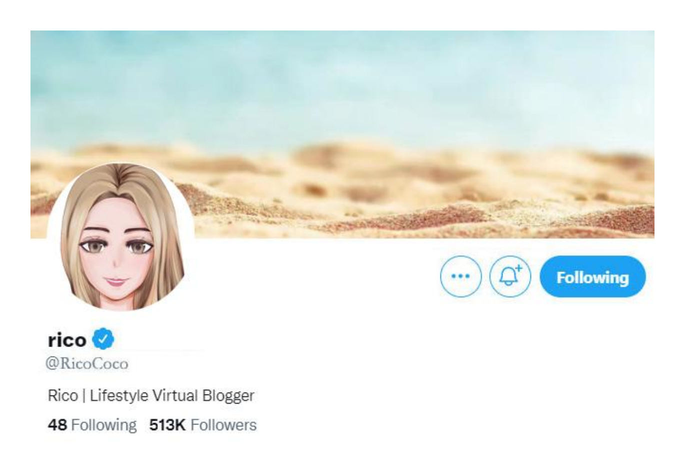
1. Rico 在道德上应该为她偷税漏税的行为负责。
2. Rico 应该为她偷税漏税的行为受到处罚。
3. 偷税漏税的行为不应该归咎于 Rico。
4. 偷税漏税的责任不应该由 Rico 承担。

[操纵检验]

就你而言，Rico 是？(0 = 真人博主, 1 = 虚拟博主)

英国实验材料

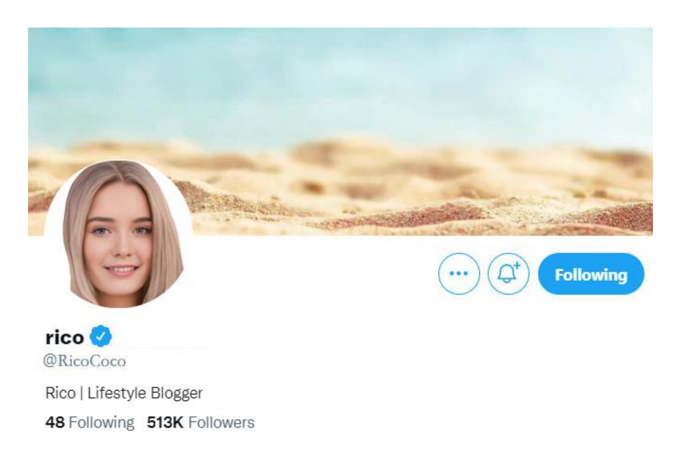
[虚拟人组操纵材料]



Rico is a virtual blogger on Twitter. She actively shares her outfit and lifestyle on social media and she also recommends nice products through livestreaming on social media every week.

(Note: a virtual blogger is a digital character created in computer graphics software, then given a personality defined by a first−person view of the world, and made accessible on media platforms.)

[真人组操纵材料]



Rico is a blogger on Twitter. She actively shares her outfit and lifestyle on social media and she also recommends nice products through livestreaming on social media every week.

[不道德行为情景]

Recently, Rico was suspected of tax evasion on her income in livestreaming e-commerce. Now, she is under investigation.

[道德责任测量]

To what extent do you agree with the following statements? (1 = completely disagree, 7 = completely agree)

1. Rico is morally responsible for her tax evasion.
2. Rico should be punished for her tax evasion.
3. Rico should not be blamed for her tax evasion.
4. Rico should not be held accountable for her tax evasion.

[操纵检验]

In your opinion, Rico is?

0 = a human blogger

1 = a virtual blogger

实验 1c 操纵和主要测量材料

<u>中国实验材料</u>

[虚拟人工智能组操纵材料]



Rico 是微博上的一个由人工智能和算法驱动的虚拟博主。她经常活跃在社交媒体上分享她的穿搭和生活。

（注：虚拟博主是利用计算机图像软件制作的数字形象，他们以第一人称视角看待世界，并在媒体平台上占据一席之地。）

[虚拟真人组操纵材料]



Rico 是微博上的一个由人工智能和算法驱动的虚拟博主。她经常活跃在社交媒体上分享她的穿搭和生活。

（注：虚拟博主是利用计算机图像软件制作的数字形象，他们以第一人称视角看待世界，并在媒体平台上占据一席之地。）

[真人组操纵材料]



Rico 是微博上的一个博主。她经常活跃在社交媒体上分享她的穿搭和生活。

[不道德行为情景]

最近，Rico 在微博上公开发表的一张图片被指控侵犯了某位摄影师的作品版权。事后，Rico 删除了这张图片并公开道歉。

[道德责任测量(选自 Gray & Wegner, 2009)]

1.你认为对于侵犯版权的行为， Rico 应该承担多少责任？(1 = 没有责任, 7 = 很大责任)

2.你认为对于侵犯版权的行为， 多大程度上应该归咎于 Rico？(1 = 不应该, 7 = 非常应该)

[虚拟人背后主体操纵检验]

就你而言， 驱动 Rico 的背后主体是(0 = 真实人类， 1 = 人工智能)

[主观社会经济地位测量(Adler et al. 2000)]

把这个梯子想象成代表你的国家人民所处的位置。

在这个阶梯的顶端是那些最富有的人——那些拥有最多的钱、最多的教育和最受尊敬的工作的人。

最底层的人境况最差——他们的钱最少, 受教育程度最低, 工作最不受尊重, 或者没有工作。

你在这个阶梯上的位置越高， 你就越接近最高层的人； 你越低， 你就越接近最底层的人。



你会把自己放在这个梯子上的什么位置？请选择下面的数字(从 1 到 10), 与中国其他人相比, 你认为自己此时所处的位置。

[受教育程度测量]

请选择您的最高学历

1 = 初中及以下

2 = 普高/中专/技校/职高

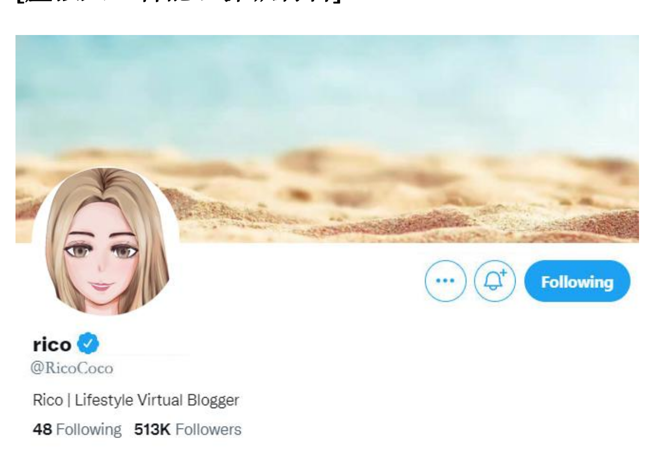
3 = 专科及本科

4 = 硕士

5 = 博士及以上

英国实验材料

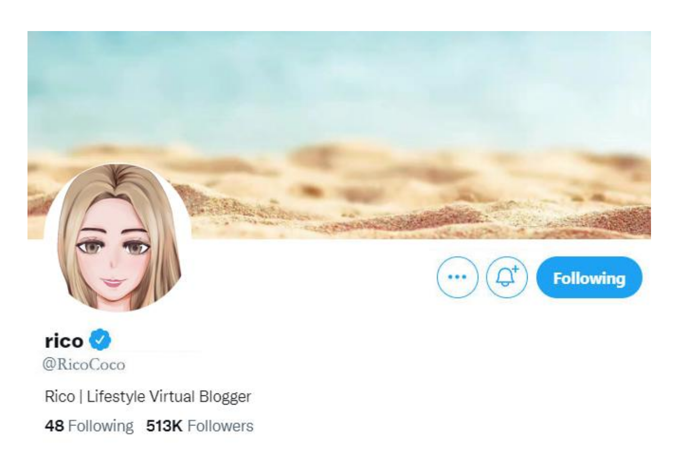
[虚拟人工智能组操纵材料]



Rico is a virtual blogger powered by artificial intelligence and algorithm on Twitter. She actively shares her outfit and lifestyle on social media.

(Note: virtual blogger is a digital character created in computer graphics software, then given a personality defined by a first-person view of the world, and made accessible on media platforms.)

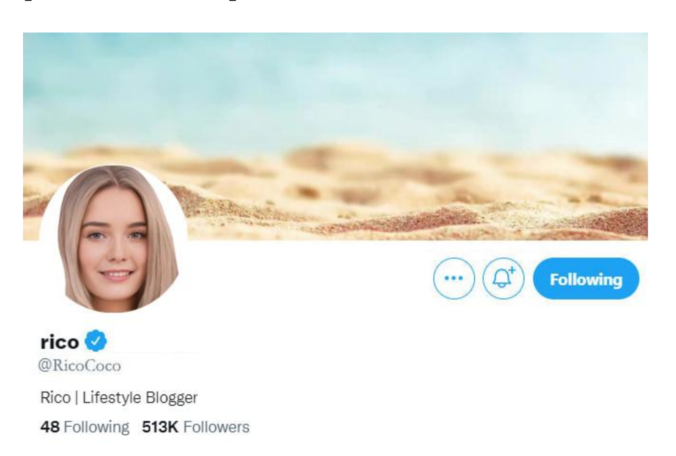
[虚拟真人组操纵材料]



Rico is a virtual blogger powered by real humans on Twitter. She actively shares her outfit and lifestyle on social media.

(Note: a virtual blogger is a digital character created in computer graphics software, then given a personality defined by a first-person view of the world, and made accessible on media platforms.)

[真人组操纵材料]



Rico is a blogger on Twitter. She actively shares her outfit and lifestyle on social media.

[不道德行为情景]

Recently, a photo posted by Rico on Twitter is accused of infringing the copyright of a photographer's work. Afterwards, Rico deleted the photo and publicly apologized.

[道德责任测量]

1. How responsible do you think Rico is for copyright infringement? (1 = not at all, 7 = very much)
2. How much blame do you think Rico deserves for copyright infringement? (1 = not at all, 7 =

very much)

[虚拟人背后主体操纵检验]

In your opinion, Rico is powered by?

0 = real humans

1 = artificial intelligence

[主观社会经济地位测量(Adler et al. 2000**)]**

Think of this ladder as representing where people stand in your homeland.

At the top of the ladder are the people who are the best off – those who have the most money, the most education and the most respected jobs.

At the bottom are the people who are the worst off–who have the least money, the least education and the least respected jobs or no jobs.

The higher you are on this ladder, the closer you are to the people at the very top; the lower you are, the closer you are to the people at the very bottom.



Where would you place yourself on this ladder? Please choose the number of the rung below (ranging from 1−10), where you think you stand at this time in your life, relative to other people in the United States.

[受教育程度测量]

Please indicate your highest level of education.

1 = Lower degree

2 = High school degree

3 = College degree

4 = Master’s degree

5 = Higher degree, including doctorate and law degree

实验 2 主要测量材料

中国实验材料

[虚拟人组操纵材料]

同实验 1b。

[真人组操纵材料]

同实验 1b。

[不道德行为情景]

最近，Rico 被曝出在直播带货的收入上存在偷税漏税的行为。现在，她正在被调查中。

[道德责任测量(选自 Gray & Wegner, 2009)]

1.你认为对于偷税漏税的行为，Rico 应该承担多少责任？(1= 没有责任, 7 = 很大责任)

2.你认为对于偷税漏税的行为，Rico 应该承担多少指责？(1 = 没有指责, 7 = 很多指责)

[感知心智能力测量(Gray et al., 2011)]

Rico 在多大程度上具备以下能力？

	1 = 没有能力	2	3	4	5	6	7 = 很有能力
自控	○	○	○	○	○	○	○
举止道德	○	○	○	○	○	○	○
计划	○	○	○	○	○	○	○
沟通	○	○	○	○	○	○	○
记忆	○	○	○	○	○	○	○
思考	○	○	○	○	○	○	○
感到快乐	○	○	○	○	○	○	○
感到渴望	○	○	○	○	○	○	○
感到害怕	○	○	○	○	○	○	○
感到痛苦	○	○	○	○	○	○	○
感到愤怒	○	○	○	○	○	○	○
感到开心	○	○	○	○	○	○	○

[对虚拟人熟悉度测量(Carlson et al., 2020)]

请选择合适的回答来描述你对虚拟博主的了解，如初音未来，洛天依，微软小冰等。

(注：虚拟博主是利用计算机图像软件制作的数字形象，他们以第一人称视角看待世界，并在媒体平台上占据一席之地。)

	1	2	3	4	5	6	7	
非常不熟悉	○	○	○	○	○	○	○	非常熟悉
完全不知情	○	○	○	○	○	○	○	非常知情
什么都不了解	○	○	○	○	○	○	○	很了解

[不道德行为严重性测量]

你认为偷税漏税的行为有多严重？(1 = 完全不严重, 7 = 非常严重)

美国实验材料

[虚拟人组操纵材料]

同实验 1b。

[真人组操纵材料]

同实验 1b。

[不道德行为情景]

Recently, Rico was suspected of tax evasion on her income in livestreaming e−commerce. Now, she is under investigation.

[道德责任测量(选自 Gray & Wegner, 2009)]

1. How responsible do you think Rico is for tax evasion? (1 = not at all, 7 = very much)
2. How much blame do you think Rico deserves for tax evasion? (1 = not at all, 7 = very much)

[感知心智能力测量(Gray et al., 2011)]

How much is Rico capable of the following abilities?

	1 = not capable at all	2	3	4	5	6	7 = very capable
self−control	○	○	○	○	○	○	○
acting morally	○	○	○	○	○	○	○
planning	○	○	○	○	○	○	○
communication	○	○	○	○	○	○	○
memory	○	○	○	○	○	○	○
thought	○	○	○	○	○	○	○
feeling pleasure	○	○	○	○	○	○	○
feeling desire	○	○	○	○	○	○	○
feeling fear	○	○	○	○	○	○	○
feeling pain	○	○	○	○	○	○	○

feeling rage	○	○	○	○	○	○	○
feeling joy	○	○	○	○	○	○	○

[对虚拟人熟悉度测量(Carlson et al., 2020)]

Please select the appropriate response that best describes your knowledge of virtual bloggers, such as Hatsune Miku, Kizuna AI, Lil Miquela and so on.

(Note: a virtual blogger is a digital character created in computer graphics software, then given a personality defined by a first−person view of the world, and made accessible on media platforms.)

	1	2	3	4	5	6	7	
very unfamiliar	○	○	○	○	○	○	○	very familiar
not at all informed	○	○	○	○	○	○	○	highly informed
know nothing at all	○	○	○	○	○	○	○	know a great deal

[不道德行为严重性测量]

How serious do you think the behavior of tax evasion is? (1 = not at all, 7 = very serious)

实验 3 操纵和主要测量材料

<u>中国实验材料</u>

[虚拟人实验材料]



Rico，又名 Coco，出生于 2002 年 3 月 21 日。她于 2021 年以虚拟唱跳偶像的身份出道并且吸引了很多粉丝。她的穿搭别具一格，走在时尚前线。Rico 经常在社交媒体上分享她的日常生活。

（注：虚拟偶像是利用计算机图像软件制作的数字形象，他们以第一人称视角看待世界，并在媒体平台上占据一席之地。）

[不道德行为情景]



rico
22-3-7 10:44
快来听我的新原创专辑#Youthstory#
https://video/BV1Jy4y157Qy/

Rico 在今年三月发布了一首原创单曲。然而，这张单曲被曝出抄袭了另一位歌手的歌曲。事后，Rico 为此事公开道歉。

[道德惩罚测量]

1. 罚款：你认为 Rico 抄袭作品在多大程度上应该被罚款？(1 = 不应该, 7 = 非常应该)
2. 封禁账号：你认为 Rico 的账号是否应该被封禁？(1 = 不应该, 7 = 非常应该)
3. 取消关注：假设你已经关注了 Rico 一段时间，对于 Rico 抄袭作品，你是否会取消关注？(0 = 继续关注, 1 = 取消关注)

[道德责任测量(选自 Gray & Wegner, 2009)]

1. 你认为对于抄袭的行为，Rico 应该承担多少责任？(1 = 没有责任, 7 = 很大责任)
2. 你认为对于抄袭的行为，Rico 应该承担多少指责？(1 = 没有指责, 7 = 很多指责)

[感知心智能力测量(Gray et al., 2011)**]**

同实验 2。

[主观社会经济地位测量(Adler et al. 2000)**]**

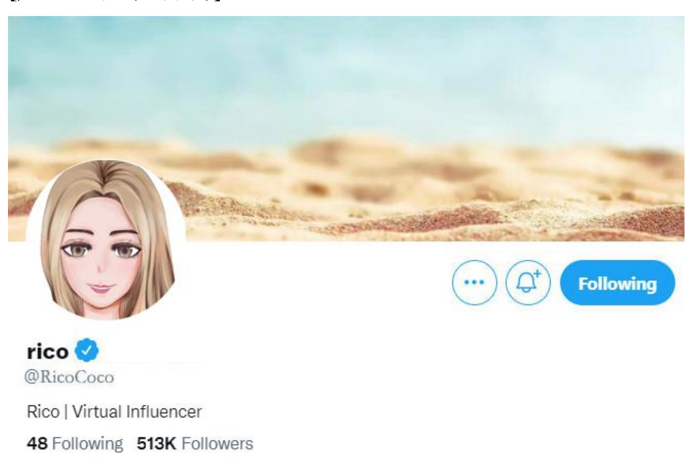
同实验 1c。

[受教育程度测量]

同实验 1c。

<u>美国实验材料</u>

[虚拟人实验材料]



Rico is a virtual influencer who debuted last year, below is her detailed information.

Rico, also known as Coco, was born on March 21, 2002. She debuted as a virtual singing and dancing influencer in May 2021 and has attracted a lot of fans. She has her own dressing style and always follows the latest fashion. Rico often shares her daily life on social media.

(Note: a virtual influencer is a digital character created in computer graphics software, then given a personality defined by a first – person view of the world, and made accessible on media platforms.)

[不道德行为情景]



The virtual influencer Rico released her original EP in March this year. However, it was found that this EP was plagiarized from another singer. And later, Rico made a public apology.

[道德惩罚测量]

1. 罚款：To what extent do think Rico should be fined for plagiarizing? (1 = not at all, 7 = very much)
2. 封禁账号：To what extent do you think Rico's Twitter account should be suspended? (1 = not at all, 7 = very much)
3. 取消关注：Suppose you have been following the virtual influencer Rico for a while. Will you choose to unfollow Rico for her plagiarizing? (0 = continue to follow, 1 = unfollow)

[道德责任测量]

1. How responsible do you think Rico is for plagiarizing? (1 = not at all, 7 = very much)
2. How much blame do you think Rico deserves for plagiarizing? (1 = not at all, 7 = very much)

[感知心智能力测量]

同实验 2。

[主观社会经济地位测量]

同实验 1c。

[受教育程度测量]

同实验 1c。

附录二：预实验：感知心智能力的文化差异

预实验的目的是检验中西方文化对虚拟人心智能力上评价是否存在显著的差异。在本实验中，我们采用英国样本作为西方文化的代表。我们认为，比起西方文化，中国文化影响下的人们对虚拟人心智能力的评价更高。

实验分别通过国外问卷平台 Prolific 和国内问卷平台 Credamo 各邀请了 199 名英国白人被试(80.40%女性; $M_{年龄}$ = 30.62 岁, $SD_{年龄}$ = 5.80 岁)和 200 名中国被试(77.50%女性; $M_{年龄}$ = 30.89 岁, $SD_{年龄}$ = 4.95 岁)参与研究，并以小额现金作为实验报酬。在英国被试中，真人组有 104 人，虚拟人组有 95 人；在中国被试中，真人组有 100 人，虚拟人组有 100 人。

与正式实验 1a 的操纵方式和内容一致，在真人组中，中国(英国)的被试会看到博主 Rico 的微博(Twitter)主页；在虚拟人组中，中国(英国)的被试会看到虚拟博主 Rico 的微博(Twitter)主页。接着，为了测量感知心智能力，被试分别从认知自主力(Cronbach's α = 0.88)和情绪感知力(Cronbach's α = 0.95)两个方面评价了 Rico 的心智能力(7 分量表, 1 = 没有能力, 7 = 很有能力; Gray et al., 2011; Cronbach's α = 0.93)，每一个条目的评价顺序随机呈现。最后，所有被试报告了性别和年龄。

以文化类型(英国编码为 0，中国编码为 1)和博主类型(真人编码为 0，虚拟人编码为 1)分别作为自变量，感知心智能力作为因变量进行方差分析。结果显示，文化类型的主效应显著(F(1, 395) = 3.88, p = 0.049, η_p^2 = 0.010)，博主类型的主效应显著(F(1, 395) = 48.59, p < 0.001, η_p^2 = 0.110)，文化类型和博主类型对心智能力评价的交互效应显著(F(1, 395) = 38.00, p < 0.001, η_p^2 = 0.088)。比起英国人，中国人认为虚拟人的心智能力更高($M_{中国}$ = 4.57, SD = 1.08, 95%CI = [4.34, 4.80]; $M_{英国}$ = 3.62, SD = 1.49, 95%CI = [3.39, 3.86]; F(1, 395) = 32.35, p < 0.001, η_p^2 = 0.076)；并且中国人认为虚拟人和真人的心智能力相似(F(1, 395) = 0.33, p = 0.568)。

心智能力分为认知自主力和情绪感知力两个维度，文化类型和博主类型对认知自主力(F(1, 395) = 12.26, p = 0.001, η_p^2 = 0.030)和情绪感知力(F(1, 395) = 52.25, p < 0.001, η_p^2 = 0.117)评价的交互效应都有显著。具体而言，比起英国人，中国人认为虚拟人的认知自主力($M_{中国}$ = 4.76, SD = 1.01, 95%CI = [4.52, 5.00]; $M_{英国}$ = 4.20, SD = 1.45, 95%CI = [3.95, 4.45]; F(1, 395) = 10.32, p = 0.001, η_p^2 = 0.025)和情绪感知力($M_{中国}$ = 4.38, SD = 1.33, 95%CI = [4.11, 4.66]; $M_{英国}$ = 3.04, SD = 1.93, 95%CI = [2.76, 3.33]; F(1, 395) = 44.70, p < 0.001, η_p^2 = 0.102)都更高；并且中国人认为虚拟人和真人的认知自主力(F(1, 395) = 0.51, p = 0.475)和情绪感知力(F(1, 395) = 2.49, p = 0.115)都没有显著差异。

在控制年龄和性别后，文化类型和博主类型对心智能力(F(1, 393) = 37.76, p < 0.001, η_p^2 = 0.088)评价的交互效应依然显著，包括认知自主力(F(1, 393) = 11.85, p = 0.001, η_p^2 = 0.029)和情绪感知力(F(1, 393) = 52.59, p < 0.001, η_p^2 = 0.118)。具体而言，比起英国人，中国人认为

虚拟人具有更高的心智能力($F(1, 393) = 32.43$, $p < 0.001$, $\eta_p^2 = 0.076$)，包括认知自主力($F(1, 393) = 10.22$, $p = 0.002$, $\eta_p^2 = 0.025$)和情绪感知力($F(1, 393) = 45.11$, $p < 0.001$, $\eta_p^2 = 0.103$)；并且中国人认为虚拟人和真人的心智能力($F(1, 393) = 0.32$, $p = 0.579$)、认知自主力($F(1, 393) = 0.34$, $p = 0.558$)和情绪感知力($F(1, 393) = 2.13$, $p = 0.145$)都没有显著差异。因此，控制年龄和性别并不影响实验结果的显著性。以上结果表明，相比于西方文化，中国文化影响下的人们认为虚拟人具有更高的心智能力。

附录三：采用男性实验材料复制中介实验

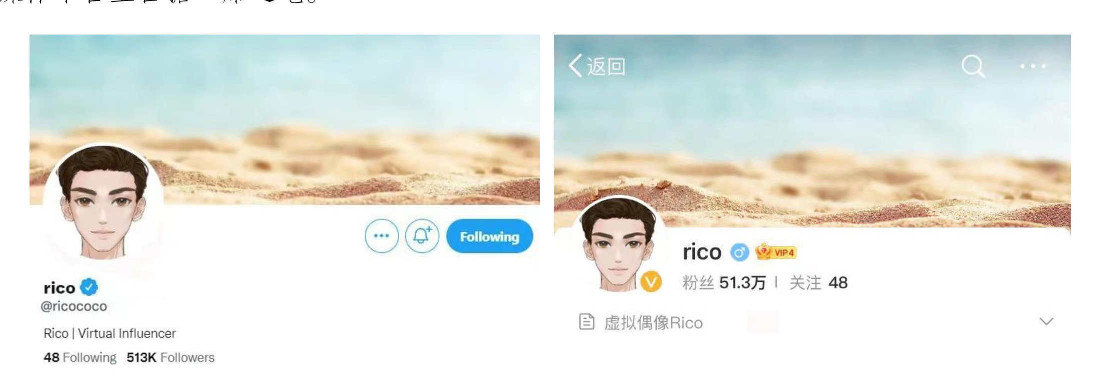
由于正文实验中所采用的虚拟人和真人的实验材料均为女性，本实验的目的是更换男性实验材料复制中介效应的实验结果。正文的实验已经表明，中西方文化差异并不会影响对真人的道德责任判断。因此，本实验聚焦于对虚拟人的道德责任判断。我们认为，采用男性实验材料并不会改变对虚拟人道德责任判断的中西方文化差异。

实验分别通过国外问卷平台 Prolific 和国内问卷平台 Credamo 各邀请了 101 名美国白人被试(49.50%女性; $M_{年龄}$ = 34.48 岁, $SD_{年龄}$ =8.58 岁)和 100 名中国被试(56.00%女性; $M_{年龄}$ = 31.49 岁, $SD_{年龄}$ = 7.65 岁)参与研究，并以小额现金作为实验报酬。

与实验 2 的操纵方式一致，不同的是，本实验中采用的是男性虚拟偶像(如下图所示)及相关描述：

“Rico 是一个 21 岁的[虚拟]偶像。Rico 在短短一年内粉丝数量就超过了十万。他在社交媒体上分享生活中发生的各种事情。他的帖子内容在评论区中引发了激烈的讨论，他还不断收到数以千计的私信，他通过回复私信与世界各地的粉丝互动。

虚拟偶像是利用计算机图像软件制作的数字形象，他们以第一人称视角看待世界，并在媒体平台上占据一席之地。”



与实验 2 一致，被试在看过有关虚拟人 Rico 的描述后被告知，“最近，Rico 被曝出在直播带货的收入上存在偷税漏税的行为。现在，他正在被调查中。”接着，本实验采用与实验 1b 相同的四个条目的道德责任量表来测量被试对 Rico 道德责任的判断(7 分量表, 1 = 完全不同意, 7 = 完全同意; Cameron et al., 2010; Cronbach’s α = 0.81)。同样，为了测量感知心智能力，与实验 2 一样，被试分别从认知自主力(Cronbach’s α = 0.92)和情绪感知力(Cronbach’s α = 0.96)两个方面评价了 Rico 的心智能力(7 分量表, 1 = 没有能力, 7 = 很有能力; Gray et al., 2011; Cronbach’s α = 0.94)。最后，所有被试报告了主观社会经济地位、受教育程度、性别和年龄。

以文化类型(美国编码为 0，中国编码为 1)为自变量，道德责任判断作为因变量进行方差分析。结果显示，比起美国人，中国人认为虚拟人应该承担更大的道德责任($M_{中国}$ = 5.46, SD = 1.24, 95%CI = [5.22, 5.71]; $M_{美国}$ = 4.93, SD = 1.56, 95%CI = [4.62, 5.24]; $F(1, 199) = 7.23$, p = 0.008, $\eta_p^2 = 0.035$)。以心智能力评价作为因变量进行方差分析，结果显示，比起美国人，中国人认为虚拟人具有更高的心智能力($M_{中国}$ = 3.97, SD = 1.39, 95%CI = [3.69, 4.24]; $M_{美国}$

$= 3.32$, $SD = 1.61$, 95%CI = [3.00, 3.64]; $F(1, 199) = 9.40$, $p = 0.002$, $\eta_p^2 = 0.045$)，包括认知自主力($M_{中国} = 4.12$, $SD = 1.50$, 95%CI = [3.82, 4.42]; $M_{美国} = 3.68$, $SD = 1.52$, 95%CI = [3.38, 3.98]; $F(1, 199) = 4.20$, $p = 0.042$, $\eta_p^2 = 0.021$)和情绪感知力($M_{中国} = 3.82$, $SD = 1.45$, 95%CI = [3.53, 4.10]; $M_{美国} = 2.95$, $SD = 2.01$, 95%CI = [2.56, 3.35]; $F(1, 199) = 12.26$, $p = 0.001$, $\eta_p^2 = 0.058$)。

以文化类型(美国编码为 0，中国编码为 1)为自变量，感知心智能力为中介变量，道德责任判断为因变量，在样本选择为 5000 次、95%的置信区间下采用 Bootstrapping (PROCESS Model 4; Hayes, 2015)分析中介效应。结果显示，感知心智能力的中介效应成立(indirect effect $\beta = 0.16$, $SE = 0.06$, 95%CI = [0.05, 0.29])。同时将认知自主力和情绪感知力两个维度作为中介变量进行平行中介分析，结果显示，情绪感知力的中介效应成立(indirect effect $\beta = 0.31$, $SE = 0.12$, 95%CI = [0.11, 0.57])但认知自主力的中介效应不成立(indirect effect $\beta = -0.07$, $SE = 0.06$, 95%CI = [−0.20, 0.01])。这些结果表明，心智能力中的情绪感知力而非认知自主力是影响中西方人对虚拟人道德责任判断的主要原因，复制了实验 2 的结果。

在控制了主观社会经济地位、受教育程度、性别和年龄后，文化类型对道德责任判断($F(1, 195) = 8.79$, $p = 0.003$, $\eta_p^2 = 0.043$)和心智能力评价($F(1, 195) = 12.59$, $p < 0.001$, $\eta_p^2 = 0.061$)的主效应依然显著，包括认知自主力($F(1, 195) = 4.80$, $p = 0.030$, $\eta_p^2 = 0.024$)和情绪感知力($F(1, 195) = 17.75$, $p < 0.001$, $\eta_p^2 = 0.083$)。感知心智能力的中介效应依然成立(indirect effect $\beta = 0.19$, $SE = 0.07$, 95%CI = [0.07, 0.34])，并且平行中介分析结果表明，主要由情绪感知力(indirect effect $\beta = 0.37$, $SE = 0.12$, 95%CI = [0.16, 0.63])而非认知自主力(indirect effect $\beta = -0.07$, $SE = 0.06$, 95%CI = [−0.21, 0.02])中介。因此，控制主观社会经济地位、受教育程度、性别和年龄并不影响实验结果的显著性。

附录四：加入控制变量之后的结果

实验 1a 控制年龄和性别之后的数据结果

以文化类型(美国编码为 0，中国编码为 1)为自变量，博主类型(真人编码为 0，虚拟人编码为 1)为调节变量，道德责任判断为因变量，性别和年龄为控制变量进行回归分析，文化类型和博主类型的交互效应的结果依然显著($\beta = 0.61$, $SE = 0.25$, $p = 0.017$, 95%CI = [0.11, 1.11])；在真人组，中西方文化差异不显著($\beta = 0.07$, $SE = 0.19$, $p = 0.725$, 95%CI = [−0.30, 0.43])；在虚拟人组，中国人认为虚拟人需要承担更大的道德责任($\beta = 0.68$, $SE = 0.19$, $p < 0.001$, 95%CI = [0.31, 1.04])；在中国被试组，虚拟人和真人在道德责任的判断上不存在显著差异($\beta = -0.03$, $SE = 0.18$, $p = 0.858$, 95%CI = [−0.39, 0.32])，因此，与原先结果一致。

实验 1b 控制年龄和性别之后的数据结果

以文化类型(英国编码为 0，中国编码为 1)为自变量，博主类型(真人编码为 0，虚拟人编码为 1)为调节变量，道德责任判断为因变量，性别和年龄为控制变量进行回归分析，文化类型和博主类型的交互效应的结果依然显著($\beta = 0.80$, $SE = 0.25$, $p = 0.002$, 95%CI = [0.31, 1.30])；在真人组，中西方文化差异不显著($\beta = -0.004$, $SE = 0.18$, $p = 0.982$, 95%CI = [−0.36, 0.35])；在虚拟人组，中国人认为虚拟人需要承担更大的道德责任($\beta = 0.80$, $SE = 0.18$, $p < 0.001$, 95%CI = [0.44, 1.16])；在中国被试组，虚拟人和真人在道德责任的判断上不存在显著差异($\beta = -0.28$, $SE = 0.18$, $p = 0.121$, 95%CI = [−0.63, 0.07])，因此，与原先结果一致。

实验 1c 控制年龄、性别、受教育程度和主观社会经济地位之后的数据结果

以文化类型(英国编码为 0，中国编码为 1)、博主类型(真人编码为 0，虚拟人编码为 1)、性别和受教育程度为自变量，以年龄和社会经济地位为协变量，以道德责任判断为因变量进行方差分析，文化类型和博主类型的交互效应的结果依然显著($F(2, 587) = 12.13$, $p < 0.001$, $\eta_p^2 = 0.040$)；在真人组，中西方文化差异不显著($F(1, 587) = 2.68$, $p = 0.102$)；在虚拟真人组，中国人认为虚拟人应该承担更大的道德责任($F(1, 587) = 64.83$, $p < 0.001$, $\eta_p^2 = 0.099$)；在虚拟人工智能组，中国人认为虚拟人应该承担更大的道德责任($F(1, 587) = 42.94$, $p < 0.001$, $\eta_p^2 = 0.068$)；在中国被试组，两种类型的虚拟人和真人在道德责任的判断上不存在显著差异($F(2, 587) = 1.77$, $p = 0.171$)，两两比较的结果也均不显著($ps > 0.221$)。因此，与原先结果一致。

实验 2 控制年龄和性别之后的数据结果

以文化类型(美国编码为 0，中国编码为 1)为自变量，博主类型(真人编码为 0，虚拟人编码为 1)为调节变量，以性别和年龄为控制变量，以道德责任判断为因变量进行回归分析，文化类型和博主类型的交互效应结果显著($\beta = 0.94$, $SE = 0.28$, $p < 0.001$, 95%CI = [0.39, 1.48])；在真人组，中西方文化差异不显著($\beta = 0.28$, $SE = 0.20$, $p = 0.162$, 95%CI = [−0.11,

0.67])；在虚拟人组，中国人认为虚拟人需要承担更大的道德责任($\beta = 1.22$, $SE = 0.20$, $p < 0.001$, 95%CI = [0.83, 1.61])。

以感知心智能力为因变量进行回归分析，文化类型和博主类型的交互效应结果显著($\beta = 0.98$, $SE = 0.22$, $p < 0.001$, 95%CI = [0.54, 1.42])；在真人组，中西方文化差异不显著($\beta = -0.15$, $SE = 0.16$, $p = 0.357$, 95%CI = [−0.46, 0.17])；在虚拟人组，中国人认为虚拟人的心智能力显著更高($\beta = 0.83$, $SE = 0.16$, $p < 0.001$, 95%CI = [0.52, 1.14])。将心智能力划分为认知自主力和情绪感知力后，以认知自主力为因变量进行回归分析，文化类型和博主类型的交互效应结果显著($\beta = 0.74$, $SE = 0.23$, $p = 0.002$, 95%CI = [0.27, 1.20])；在真人组，中西方文化差异不显著($\beta = -0.08$, $SE = 0.17$, $p = 0.618$, 95%CI = [−0.42, 0.25])；在虚拟人组，中国人认为虚拟人的认知自主力显著更高($\beta = 0.65$, $SE = 0.17$, $p < 0.001$, 95%CI = [0.32, 0.98])。以情绪感知力为因变量，文化类型和博主类型的交互效应结果显著($\beta = 1.22$, $SE = 0.27$, $p < 0.001$, 95%CI = [0.70, 1.75])；在真人组，中西方文化差异不显著($\beta = -0.21$, $SE = 0.19$, $p = 0.274$, 95%CI = [−0.59, 0.17])；在虚拟人组，中国人认为虚拟人的情绪感知力显著更高($\beta = 1.01$, $SE = 0.19$, $p < 0.001$, 95%CI = [0.64, 1.39])。

感知心智能力有调节的中介效应依旧成立(indirect effect = 0.35, $SE = 0.11$, 95%CI = [0.16, 0.59])；在真人组，感知心智能力的中介作用不成立(indirect effect = −0.05, $SE = 0.05$, 95%CI = [−0.17, 0.45])；在虚拟人组中，感知心智能力的中介作用成立(indirect effect = 0.29, $SE = 0.09$, 95%CI = [0.14, 0.48])。以认知自主力和情绪感知力作为中介变量作平行中介分析，情绪感知力有调节的中介效应依然成立(indirect effect = 0.47, $SE = 0.14$, 95%CI = [0.22, 0.78])；在真人组，情绪感知力的中介作用不成立(indirect effect = −0.08, $SE = 0.07$, 95%CI = [−0.22, 0.04])；在虚拟人组中，情绪感知力的中介作用成立(indirect effect = 0.39, $SE = 0.12$, 95%CI = [0.19, 0.64])；认知自主力有调节的中介效应不成立(indirect effect = −0.05, $SE = 0.06$, 95%CI = [−0.19, 0.07])。以上结果的显著性及方向均和原先结果一致。

实验 3 控制年龄、性别、受教育程度和主观社会经济地位之后的数据结果

以文化类型(美国编码为 0，中国编码为 1)、性别和受教育程度为自变量，以年龄和社会经济地位为协变量，以“罚款”($F(1, 192) = 29.90$, $p < 0.001$, $\eta_p^2 = 0.135$)、道德责任判断($F(1, 192) = 10.52$, $p = 0.001$, $\eta_p^2 = 0.052$)、感知心智能力($F(1, 192) = 35.16$, $p < 0.001$, $\eta_p^2 = 0.155$；认知自主力: $F(1, 192) = 11.46$, $p = 0.001$, $\eta_p^2 = 0.056$；情绪感知力: $F(1, 192) = 44.63$, $p < 0.001$, $\eta_p^2 = 0.189$)分别为因变量进行方差分析，文化类型的主效应均显著；但以“封禁账号”为因变量进行方差分析，文化类型的主效应边际显著($F(1, 192) = 3.47$, $p = 0.064$, $\eta_p^2 = 0.018$)。以“取消关注”为因变量，文化类型、性别和受教育程度为分类自变量，年龄和社会经济地位为协变量进行方差分析，文化类型的主效应也显著($B = 0.83$, $SE = 0.34$, Wald $\chi^2 = 5.93$, $p = 0.015$, Exp(B) = 2.30)。因此，与原先结果基本一致。